UNIVERSAL
LIBRARY
OU_172002
UNIVERSAL
LIBRARY

# PHILOSOPHICAL

# TRANSACTIONS

OF THE

# ROYAL SOCIETY OF LONDON.

## SERIES A.

CONTAINING PAPERS OF A MATHEMATICAL OR PHYSICAL CHARACTER.

VOL. 221.

PRINTED AND PUBLISHED FOR THE ROYAL SOCIETY BY
HARRISON AND SONS, LTD., 44-47, ST MARTIN'S LANE, LONDON, W.C. 2.
Printers in Ordinary to His Majesty.

MARCH, 1921.

# CONTENTS.

## (A)

## VOL. 221.

List of Illustrations . . . . . . page v
Advertisement . . . . . . . . vii

I. *Some Measurements of Atmospheric Turbulence. By* LEWIS F. RICHARDSON. *Communicated by Sir* NAPIER SHAW, *F.R.S.* . . . . . . . *page* 1

II. *On a Theory of the Second Order Longitudinal Spherical Aberration for a Symmetrical Optical System. By* T. Y. BAKER, *B.A., Instructor Commander, R.N., and* L. N. G. FILON, *M.A., D.Sc., F.R.S., Goldsmid Professor of Applied Mathematics and Mechanics in the University of London* . . 29

III. *Investigations on Lightning Discharges and on the Electric Field of Thunderstorms. By* C. T. R. WILSON, *F.R.S., Reader in Electrical Meteorology in the University of Cambridge and Observer in Meteorological Physics at the Solar Physics Observatory, Cambridge. (Communicated by the Director of the Solar Physics Observatory)* . . . . . . . . . . . . . 73

IV. *Researches on the Elastic Properties and the Plastic Extension of Metals. By* W. E. DALBY, *F.R.S., Professor of Engineering at the City and Guilds (Engineering) College of the Imperial College of Science and Technology* . 117

V. *The Stress-strain Properties of Nitro-cellulose and the Law of its Optical Behaviour. By Prof.* E. G. COKER, *F.R.S., and* K. C. CHAKKO, *M.Sc., University College, London* . . . . . . . . . . . . . . 139

VI. *The Phenomena of Rupture and Flow in Solids. By* A. A. GRIFFITH, *M. Eng. (of the Royal Aircraft Establishment). Communicated by* G. I. TAYLOR, *F.R.S.* . . . . . . . . . . . . . . . . . . . . . *page* 163

VII. *Reduction of Error by Linear Compounding. By* W. F. SHEPPARD, *Sc.D., LL.M. Communicated by* E. T. WHITTAKER, *F.R.S.* . . . . . . . . 199

VIII. *Tidal Friction in Shallow Seas. By* HAROLD JEFFREYS, *M.A., D.Sc., Fellow of St. John's College, Cambridge. Communicated by Sir* NAPIER SHAW, *F.R.S.* . . . . . . . . . . . . . . . . . . . 239

IX. *Plane Stress and Plane Strain in Bipolar Co-ordinates. By* G. B. JEFFERY, *M.A., D.Sc., Fellow of University College, London. Communicated by Prof.* L. N. G. FILON, *F.R.S.* . . . . . . . . . . . . . . . 265

X. *The Aerodynamics of a Spinning Shell. By* R. H. FOWLER, E. G. GALLOP, C. N. H. LOCK *and* H. W. RICHMOND, *F.R.S.* . . . . . . . . . . 295

XI. *A Selective Hot-Wire Microphone. By* W. S. TUCKER, *D.Sc., A.R.C.Sc., and* E. T. PARIS, *M.Sc. (Lond.). Communicated by Prof.* H. L. CALLENDAR, *F.R.S.* . . . . . . . . . . . . . . . . . . . . . . 389

*Index to Volume.* . . . . . . . . . . . . . . . . . . . . . . 431

# LIST OF ILLUSTRATIONS.

Plate 1.—Mr. LEWIS F. RICHARDSON on Some Measurements of Atmospheric Turbulence.

Plates 2–5.—Mr. C. T. R. WILSON: Investigations on Lightning Discharges and on the Electric Field of Thunderstorms.

# ADVERTISEMENT

THE Committee appointed by the *Royal Society* to direct the publication of the *Philosophical Transactions* take this opportunity to acquaint the public that it fully appears, as well from the Council-books and Journals of the Society as from repeated declarations which have been made in several former *Transactions*, that the printing of them was always, from time to time, the single act of the respective Secretaries till the Forty-seventh volume; the Society, as a Body, never interesting themselves any further in their publication than by occasionally recommending the revival of them to some of their Secretaries, when from the particular circumstances of their affairs, the *Transactions* had happened for any length of time to be intermitted. And this seems principally to have been done with a view to satisfy the public that their usual meetings were then continued, for the improvement of knowledge and benefit of mankind: the great ends of their first institution by the Royal Charters, and which they have ever since steadily pursued.

But the Society being of late years greatly enlarged, and their communications more numerous, it was thought advisable that a Committee of their members should be appointed to reconsider the papers read before them, and select out of them such as they should judge most proper for publication in the future *Transactions*; which was accordingly done upon the 26th of March, 1752. And the grounds of their choice are, and will continue to be, the importance and singularity of the subjects, or the advantageous manner of treating them: without pretending to answer for the certainty of the facts, or propriety of the reasonings contained in the several papers so published, which must still rest on the credit or judgment of their respective authors.

It is likewise necessary on this occasion to remark, that it is an established rule of the Society, to which they will always adhere, never to give their opinion, as a Body, upon any subject, either of Nature or Art, that comes before them. And therefore the thanks, which are frequently proposed from the Chair, to be given to the authors of such papers as are read at their accustomed meetings, or to the persons through whose hands they received them, are to be considered in no other light than as a matter of civility, in return for the respect shown to the Society by those communications. The like also is to be said with regard to the several projects. inventions, and curiosities of various kinds, which are often exhibited to the Society; the authors whereof, or those who exhibit them, frequently take the liberty to report, and even to certify in the public newspapers, that they have met with the highest applause and approbation. And therefore it is hoped that no regard will hereafter be paid to such reports and public notices; which in some instances have been too lightly credited, to the dishonour of the Society.

# PHILOSOPHICAL TRANSACTIONS.

## I. *Some Measurements of Atmospheric Turbulence.*

*By* LEWIS F. RICHARDSON.

*Communicated by Sir* NAPIER SHAW, *F.R.S.*

Received October 16, 1919,—Read February 26, 1920.

[PLATE 1.]

CONTENTS.

Page

I. Notation . . . . . . . . . . . . . . . . . . . . . . . . . . . 1

II. Shearing stress deduced from pilot balloon observations . . . . . . . . . . . . 3

III. Eddy-viscosity from the same. It varies with height and with direction (anisotropic) . . 5

IV. Eddy-diffusivity from the dispersal of smoke or of parachutes . . . . . . . . . . 5

V. General theory of eddy-diffusivity deduced from scattering . . . . . . . . . . 8

VI. OSBORNE REYNOLDS' eddy-stresses. A measure of gustiness . . . . . . . . . . . 10

VII. Summary of theory of scattering of particles of air . . . . . . . . . . . . 15

VIII. Numerical values from scattering of particles . . . . . . . . . . . . . 19

IX. Cumulus eddies in calm weather . . . . . . . . . . . . . . . . . . 26

X. Summary . . . . . . . . . . . . . . . . . . . . . . . . . . 28

### I. NOTATION.

THE following notation is used throughout. The co-ordinate axes are a right-handed rectangular system $x$, $y$, $h$, in which $0h$ is directed vertically upwards, and $0x$ lies in any azimuth which happens to be convenient. Elements of distance to east and to north are denoted by $de$, $dn$, so that they are special cases of $dx$, $dy$. The atmospheric density is $\rho$, the pressure is $p$, acceleration of gravity is $g$, latitude $\phi$ is reckoned negative in the southern hemisphere, and $\omega$ is the angular velocity of the earth. Velocities are denoted by $v$ with a suffix to indicate the direction *towards* which they blow. Momenta per unit volume are denoted by $m_X$, $m_Y$, $m_H$. The eddy-diffusivity is denoted by a capital K as in G. I. TAYLOR's recent papers. Another, and in the author's opinion a better, measure of turbulence is $\xi$ discussed in a previous paper.* The relation K to $\xi$ is given by

$$\frac{\partial}{\partial p}\left(\xi\frac{\partial\chi}{\partial p}\right)=\frac{\partial\chi}{\partial t}=\mathrm{K}\frac{\partial^2\chi}{\partial h^2}\quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (1)$$

where $\chi$ is either potential temperature, or else mass of water or smoke per mass of atmosphere. If $\rho$ and $\xi$ were independent of height, then from (1) we should have

$$\xi=g^2\rho^2\mathrm{K}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2)$$

---

* L. F. RICHARDSON, 'Roy. Soc. Proc.,' A, vol. 96 (1919), pp. 9 to 13.

[Published June 9, 1920.

It is suggested that $\xi$ might be named "the turbulivity." Its dimensions are: $(\text{mass})^2 \times (\text{length})^{-2} \times (\text{time})^{-5}$.

The advantage of using $\xi$ instead of K is that the former enables one to allow for variations of density and of turbulence in a simple and natural manner. The disadvantage of $\xi$ is that it has no name derivable from indoor physics. It is suited to the free atmosphere. We might compromise by using in place of K or $\xi$ the "eddy-conductivity," $c$, defined by the equation

$$\frac{\partial(\rho\chi)}{\partial t} = \frac{\partial}{\partial h}\left(c\frac{\partial\chi}{\partial h}\right) \text{ or, approximately, } \frac{\partial\chi}{\partial t} = \frac{1}{\rho}\frac{\partial}{\partial h}\left(c\frac{\partial\chi}{\partial h}\right). \quad . \quad . \quad . \quad (3)$$

In so doing we gain an acceptable name "conductivity," but we lose by the explicit appearance of density in the equation. Either $c$ or $\xi$ allows variations of turbulence with height to be treated correctly, while K does not do so, as has been pointed out elsewhere by the author.* The dimensions of $c$ are $(\text{mass}) \times (\text{length})^{-1} \times (\text{time})^{-1}$.

This $c$ is of the same dimensions as the measure of turbulence discussed by W. SCHMIDT, of Vienna, under the name of "Austausch" in two important papers. ('Sitz. Akad. Wiss.,' Wien, 1917 and 1918.)

However much turbulence and density may vary with height

$$g^2\rho c = \xi. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

On the contrary if there are no variations with height,

$$c = \rho\mathrm{K}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

The six components of stress are denoted by $\widehat{xx}$, $\widehat{yy}$, $\widehat{hh}$, $\widehat{xy}$, $\widehat{yh}$, $\widehat{hx}$, as in the writings of K. PEARSON.

The convention adopted for the signs of eddy-stresses conforms to that of LOVE'S "Theory of Elasticity." Tractions are reckoned positive. That is to say, a direct stress such as $\widehat{xx}$ is positive if it be a tension, negative if a pressure; and a shearing stress such as $\widehat{xh}$ is positive when the air on that side of a level surface for which $h$ is greater (*i.e.*, above), drags the air below in the sense of $x$ increasing.

The *definition of eddy-viscosity* adopted in this paper is

$$\frac{\text{eddy shearing stress}}{\text{rate of mean shearing strain}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\mathrm{A})$$

in agreement with the definition used by W. SCHMIDT (*loc. cit.*, 1917, p. 5).

The advantage of this definition is that it is simply based on the fundamental ideas of stress and strain, as well as being in harmony with the definition adopted in the theory of viscous liquids. (*cf.*, LAMB, 'Hydrodynamics,' IV. edn., § 326).

The question may arise as to whether the viscosity defined by (A) can ever become infinite by the vanishing of the denominator. The point is discussed by the author in the paper already cited (p. 13), and the conclusion is reached that such an occurrence would be highly improbable.

---

* *Loc. cit.*

The relation of $\xi$ to the eddy-viscosity is most easily reached *via* terms $\partial m_x/\partial t$, $\partial m_Y/\partial t$ in the dynamical equations. If pressure-gradient just balanced geostrophic wind we should have

$$\frac{\partial\,(\widehat{xh})}{\partial h} = \frac{\partial m_x}{\partial t} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

Now by the definition of viscosity given above

$$\widehat{xh} = \mu \,.\, \partial v_x/\partial h. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

Substituting (7) in (6) and inserting $m_x = \rho v_x$ there results

$$\frac{\partial}{\rho\,\partial h}\left\{\mu\,\frac{\partial v_x}{\partial h}\right\} = \frac{\partial v_x}{\partial t}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

It is seen that this equation becomes identical with (1) if

$$\chi = v_x \qquad \text{and} \qquad \xi = g^2\rho\mu, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

of which the latter is the required relation.

On comparing equations (8) and (3), it is seen that eddy-viscosity, $\mu$, and eddy-conductivity, $c$, are of the same dimensions, and appear in their respective differential equations in the same way. Indeed, TAYLOR has suggested that they are equal.* This likeness would be a good argument for recording observations in terms of these two quantities instead of in terms of diffusivity K or turbulivity $\xi$.

## II. SHEARING STRESS FROM PILOT BALLOON OBSERVATIONS.

(*Condensed and revised January 22, 1920.*) EKMAN† in a remarkable paper pointed out that the total momentum of water produced by a tangential stress on the surface of the sea, in the steady state, is directed at right angles to the tangential stress, and its amount is quite independent of the value of the viscosity or of the variation of viscosity with depth. The same applies to the atmosphere. We may use this principle to find the shearing stress on the ground, provided we have a measure of what the momentum would be if the surface stress were zero.

I have taken the wind at a height of 1½ km. to 2½ km. as the standard of reference, because, by so doing, the term depending on curvature of path, and other small terms in the dynamical equations, are automatically allowed for to a first approximation. The stress at 2 km. is undoubtedly much less than that on the ground, and is neglected. It is best to select observations in which the momentum becomes nearly independent of height above 1·5 km. A table of results follows. They were computed with the help of Mrs. L. F. RICHARDSON. Dr. H. JEFFERYS says the selection will select abnormal lapse-rates and so abnormal viscosities.

---

* 'Phil. Trans.,' A, vol. 215, p. 22.

† "On the Influence of the Earth's Rotation on Ocean Currents," by V. W. EKMAN, 'Arkiv för Matem. Astr. och Fysik,' Stockholm, Bd. II., No. 11 (1905).

### TABLE I.—Shearing Stresses at the Earth's Surface.

Here $\bar{m}$ is the mean momentum per cm.$^3$ of the air between the ground and a height 2 km. above sea level.

| Place. | Latitude. | $\bar{m}$. $\frac{\text{grm.}}{\text{cm.}^2\ \text{sec.}}$. | Resultant stress $\div \bar{m}^2$. $\frac{\text{cm.}^3}{\text{grm.}}$. | Stress on ground is veered* from wind near surface. | Stress dragging ground in the direction of $\bar{m}$. dynes cm.$^{-2}$. | Stress dragging ground perpendicularly to left of $\bar{m}$. | Reference and notes. |
|---|---|---|---|---|---|---|---|
| Eskdalemuir . | °<br>55·3 | 0·72 | 6·7 | −8° | +2·9 | +1·9 | Mean of 39 observations in 1913 and 1914. Light winds omitted. |
| Lindenberg . | 52·2 | 1·24 | 1·0 | 26° | +1·5 | +0·05 | General mean. E. GOLD, Met. Office, 'Geophys. Mem.,' V., p. 143. |
| Upavon . . | 51·3 | 1·70 | 1·0 | assumed to be zero | +2·9 | +0·8 | Strong winds. DOBSON, 'Q. J. Met. Soc.,' April, 1914. |
| Batavia . . . | −6·2 | 0·26 | 1·3 | −21° | +0·08 | −0·03 | May and June . . . .<br>VAN BEMMELEN; Batavia, 'Verhandelingen,' vol. 1, 1911, pp. 13, 14. |
| | | 0·47 | 1·1 | 30° | +0·04 | −0·24 | July to September . . |
| | | 0·47 | 0·6 | ±180° | −0·12 | −0·05 | December to February . |
| The above were obtained by the method described in the preceding page. Below follows a result taken from G. I. TAYLOR's paper, 1915. | | | | | | | |
| Upavon . . . | 51·3 | as above | 0·83 | assumed to be zero | — | — | The same observations as above. |

* Veering here means turning in the sense of north to east, and this statement applies to both sides of the equator.

## III. EDDY-VISCOSITY FROM PILOT BALLOON OBSERVATIONS.

(*Abridged, January 22, 1920.*) The method of the last section will give the difference of the shearing stresses on two surfaces of any horizontal slab of air. If we choose one of the surfaces so that $\partial v/\partial h = 0$ and consequently also the stress vanishes, we obtain the stress on the other surface. This has been done for some very smooth means for Lindenberg (E. GOLD, Met. Office, 'Geophys. Mem.,' V., p. 143). The results are set out in the following table:—

TABLE II.—Lindenberg.

The $x$ axis is directed with the surface wind. The stress is that exerted by the upper on the lower layer.

| Height above mean sea, kilometres. | Eddy-shearing-stresses. | | Rates of mean shearing. | | Eddy viscosities. | |
|---|---|---|---|---|---|---|
| | $\widehat{xh}$ dynes cm.$^{-2}$. | $\widehat{yh}$ dynes cm.$^{-2}$. | $\frac{\partial v_x}{\partial h}$ sec. $^{1}$ $10^3 \times$ | $\frac{\partial v_y}{\partial h}$ sec.$^{-1}$ $10^8 \times$ | Parallel to wind. $\frac{\widehat{xh}}{\partial v_x/\partial h}$ dyne cm.$^{-2}$ sec. | Perpendicular to wind. $\frac{\widehat{yh}}{\partial v_y/\partial h}$ dyne cm.$^{-2}$ sec. |
| 1·0 | +0·05 | −0·09 | − 1·0 | − 0·2 | 50 | 450 |
| 0·8 | +0·01 | −0·37 | − 1·2 | − 1·5 | 10 | 250 |
| 0·7 | zero | | | | | |
| 0·6 | +0·02 | −0·71 | 1 | − 3·2 | 20 | 220 |
| 0·4 | +0·09 | −0·99 | 6·0 | − 9·0 | 15 | 110 |
| 0·3 | +0·39 | −1·06 | 12·5 | −12·2 | 31 | 87 |
| 0·2 | +0·69 | −0·90 | | | | |
| 0·12 ground | +1·10 | −0·64 | 21·5 ? | - 17·5 ? | 51 | 37 |

To obtain a quantity comparable with $\xi$ we must multiply the eddy-viscosity by $g^2\rho$ which is approximately 1100 c.g.s. units.

The mean of the viscosities in the two directions increases with height as we might expect from other observations (*vide* Part VIII., below, also 'Roy. Soc. Proc.,' A, vol. 96 (1919), p. 18). But the most interesting thing about this table is the *marked lack of isotropy in viscosity. The air appears to be more viscous, for large motions, across the wind than parallel to it, except just near the ground.*

## IV. EDDY-DIFFUSIVITY FROM SMOKE OR FLOATING BODIES.

Some direct measurements have been made by observing the gradually increasing scatter of smoke or other visible material carried along by the air. The changes in height of a large number of small portions of air are observed during a fixed interval of time. These changes are found to be distributed about their mean value approximately according to the ordinary "law of error." Their scatter in height is measured by the "standard deviation," computed by the familiar methods.* In order to find the diffusivity K the observations are compared with an appropriate integral of the approximate equation

$$\frac{\partial \chi}{\partial t} = \mathrm{K}\frac{\partial^2 \chi}{\partial h^2} \quad \ldots \ldots \ldots \ldots \quad (1)$$

Such a one is

$$\chi = \frac{\mathrm{A}_1}{\sqrt{4t+\mathrm{A}_2}}\, e^{\frac{-(h-\mathrm{A}_3)^2}{\mathrm{K}(4t+\mathrm{A}_2)}} \quad \ldots \ldots \ldots \ldots \quad (2)$$

where $\mathrm{A}_1$, $\mathrm{A}_2$, $\mathrm{A}_3$ are constants.

This integral represents a horizontal lamina in which the density $\chi$ is distributed about a mean height $\mathrm{A}_3$ according to the law of error. The square of the standard deviation of the mass in the lamina can be shown to be

$$(4t+\mathrm{A}_2)\frac{\mathrm{K}}{2} \quad \ldots \ldots \ldots \ldots \quad (3)$$

So if the scatter of the same set of particles be observed at the beginning and at the end of an interval T of time, it follows that

$$\mathrm{K} = \frac{1}{2\mathrm{T}}\,(\text{increase during T of square of standard deviation}) \quad \ldots \quad (4)$$

But the increase of the square of the standard deviation is equal to the square of the standard deviation of the change of height. Accordingly

$$\mathrm{K} = \frac{1}{2\mathrm{T}}\,(\text{square of standard deviation of change of height during T}) \quad . \quad . \quad (5)$$

In this last transformation we have assumed that K is sensibly independent of height. This is permissible because the range of scatter can usually be made small. For the same reason the density of the air may be taken as independent of the height, so that we may obtain from K, the constant $\xi$, which we require when pressure is taken as independent variable in place of height, in accordance with (1) above. This procedure is not perfectly satisfactory but it is very convenient. It gives $\xi = g^2\rho^2\mathrm{K}$ and "eddy-conductivity" $= \rho\mathrm{K}$.

There is no need for the changes in height to be simultaneous for all the portions of air, and in practice it is much more convenient to let them be successive.

*Varieties of particles.*—I have observed the scattering of smoke from a smouldering wick, from burning weeds, from factory chimneys and from ship's funnels: also the scattering of portions of cloud near the horizon and of puffs of ammonium chloride from a special apparatus. Chimney smoke is not to be recommended, as it rises through the air. Clouds and steam may mislead one by evaporating. The cold $NH_4Cl$ smoke proved more satisfactory in these ways. Lycopodium dust might be better still, as isolated grains would fall at a definite rate relative to the air. But I have not succeeded in making the lumps break up into grains. The downy parachute which carries the seed of the dandelion, *Taraxacum officinale*, has been found to be convenient. When the seed is broken off, the parachute falls at a rate of 10 to 15 cm. sec.$^{-1}$ in still air. The standard deviation of this rate must be allowed for. The formulæ for correction are given below. A brown parachute, twice as large each way as that of taraxacum, grows near Benson. I am indebted to the Botanical Department of the British Museum for a search among their dried specimens for a large white parachute. A splendid one came from an African plant called strophanthus. The parachute is about 6 cm. in diameter and has a long stalk by which it can be held conveniently. When the seeds were broken off the parachutes fell, in still air, at an average rate of 20 cm. sec.$^{-1}$.

---

* *Vide* 'Computer's Handbook,' M.O. 223, Section V.

Other artificial clouds, which have been used with success, are paraffin-oil vapour from an extinguished blast lamp, and smoke of phosphorus pentoxide made by dropping calcium phosphide into dilute hydrochloric acid. In strong winds the smoke from a firework known as " Vesuvius "* is convenient.

If one could mark and follow individual molecules equation (5) would give the molecular diffusivity in still air, 0·2 cm.$^2$ sec.$^{-1}$. Actually what we observe is the centre of a small puff of smoke, and this is not constantly the position of the same molecules, so that in still air we find $K = 0$. To be perfectly exact all observations of K by this method should be increased by 0·2 cm.$^2$ sec.$^{-1}$, an entirely negligible correction. In any theory the diffusivity depends on the motions which the theory does not follow in detail. In laboratory experiments, in which the molecular motion only is ignored, K is taken as 0·2 cm.$^2$ sec.$^{-1}$. In meteorological telegraphy variations of wind of less than 10 minutes' duration are ordinarily ignored, and there is an appropriate, much larger, value of the diffusivity. In a certain scheme for numerical prediction it is proposed to average the wind over periods of 6 hours, and the further variations thus omitted must be taken into account by further increase in K. It follows that the puffs of smoke should be so small as to allow the smallest eddies to be observed, and, for the last-named purpose, that the observations should be spread over a period of 6 hours. In obtaining the data in the following table I believe the former condition has been fulfilled, but the latter has not. When only the order of K or $\xi$ is required, it is enough to assume that the standard deviation is $\frac{1}{5}$ of the distance between the extremes of height observed, when the number of observations is about 40.

*Observations very near the earth's surface* have peculiarities. It is obvious that $\partial\chi/\partial h = 0$ at an impermeable horizontal surface. This condition can be satisfied in the integral by taking the portion of the distribution which would be cut off by the surface, reflecting it in the surface, and adding it to the rest of the distribution.

---

* Made by Messrs. C. T. Brock & Co.

The standard deviation is not then sufficient to give K. It is necessary to make more elaborate computations with K. PEARSON'S "incomplete normal moment functions."

These difficulties are avoided in the symmetrical case when the source of smoke is exactly on the ground, and the smoke does not rise or fall by its temperature.

The results of these measurements are set out in the last column of Table IV. It is seen there that K increases from 5 near the surface of land up to 10,000 at the height of a factory chimney.

But before considering the results further a fuller mathematical investigation will now be made.

## V. GENERAL THEORY OF EDDY-DIFFUSIVITY DEDUCED FROM SCATTERING.

The foregoing theory of the diffusion of a lamina assumes that the diffusivity is constant throughout the space, and that the density in the lamina does not vary except in the smooth regular manner indicated by the "law of error." But it is well known that the wind has an intricate structure. Thus if observations of the smoke puffs are to yield a measure of the diffusivity from the formula

$$\text{diffusivity} = \frac{\text{increase in square of standard deviation}}{\text{twice corresponding increase in time}},$$

then either the interval of time in the denominator must be long compared with the fluctuations of the wind in time, or the initial standard deviation must be large compared with the fluctuations of the wind in space, or both conditions must hold. The former condition is an inconvenient one in practice, because puffs are apt to fade before a sufficient time has passed. Dandelion parachutes, with the seeds removed, may be better than smoke for this purpose.

The following theory brings to light some of the assumptions involved in the measurement of diffusivity by smoke puffs. It was contrived specially in order to avoid "the distance through which an eddy moves before mixing with its surroundings," a quantity which occurs in TAYLOR'S theory, but which does not lend itself easily to measurement, except in the case of cumulus eddies. See Section IX. below.

The potential* temperature $\Theta$ does not change at a point moving with fluid, if radiation and precipitation can be neglected. Now let a portion of an eddy move from a height $h_1$ at time $t_1$ to a height $h_2$ at $t_2$. Then, regarding $\Theta$ as a function of $h$ and $t$, we have

$$\Theta(h_1, t_1) = \Theta(h_2, t_2). \quad \ldots \ldots \ldots \quad (1)$$

---

* Potential temperature is the temperature which the air would acquire if compressed adiabatically to a standard pressure. If $\Theta$ is to be of service in dealing with cloudy air the standard pressure must be high enough to evaporate the cloud in all samples.

From each side of (1) subtract $\Theta(h_1, t_2)$ and divide through by $t_2-t_1$. Then

$$\frac{\Theta(h_1, t_1)-\Theta(h_1, t_2)}{t_2-t_1} = \frac{\Theta(h_2, t_2)-\Theta(h_1, t_2)}{t_2-t_1} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

Now the left side of (2) is the finite difference ratio $\frac{\delta\Theta}{\delta t}$ at the height $h_1$, which is what we want to find in terms of the spacial distribution of $\Theta$. Expand the right-hand side of (2) in powers of $h_2-h_1$ by the well-known theorem in the calculus. It follows that, if subscripts indicate the time and height

$$\left(\frac{\delta\Theta}{\delta t}\right)_{h_1,\frac{t_2+t_1}{2}} = \frac{1}{t_2-t_1}\left\{\left(\frac{\partial\Theta}{\partial h}\right)_{t_2h_1}(h_2-h_1)+\left(\frac{\partial^2\Theta}{\partial h^2}\right)_{t_2h_1}\frac{(h_2-h_1)^2}{2\,!}+\text{higher terms.}\right. \quad . \quad . \quad (3)$$

The difference ratio on the left of this equation is centred at the same height $h_1$ as the differential coefficients on the right of the same, but at a time $\frac{1}{2}(t_2-t_1)$ previous. This slight misfit in centering will not matter, because $t_2-t_1$ will be of the order of one minute or less, whereas we are next going to take the average of each term in (3) over a much longer time, say 6 hours. The subscripts may now be omitted as unnecessary. Let a bar over a symbol, or group of symbols, denote the mean value over this longer period. Let a dash denote the instantaneous deviation from this mean, so that, for instance, we have for every fluctuating quantity a formula such as

$$\frac{\partial\Theta}{\partial h} = \left(\overline{\frac{\partial\Theta}{\partial h}}\right)+\left(\frac{\partial\Theta}{\partial h}\right)' \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

Now the mean of any dashed quantity vanishes. . . . . . . . . . . . (5)

Again the mean of the product of any dashed quantity into any barred quantity also vanishes. . . . . . . . . . . . . . . . . . . . . . . . (6)

We shall further suppose that the mean of $h_2-h_1$ vanishes, . . . . . . . . (7)

that is to say that there is no mean vertical displacement.

Then, in the first term on the right of (3)

$$\overline{\frac{\partial\Theta}{\partial h}(h_2-h_1)} = \overline{\left\{\left(\overline{\frac{\partial\Theta}{\partial h}}\right)+\left(\frac{\partial\Theta}{\partial h}\right)'\right\}\{(\overline{h_2-h_1})+(h_2-h_1)'\}}$$

$$= \overline{\left(\frac{\partial\Theta}{\partial h}\right)'(h_2-h_1)'} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

because of (6) and (7).

Now $\left(\frac{\partial\Theta}{\partial h}\right)'(h_2-h_1)'$, after being divided by $t_2-t_1$ and by the "standard deviations" of $\left(\frac{\partial\Theta}{\partial h}\right)'$ and of $(h_2-h_1)'$, becomes equal to the correlation between $\partial\Theta/\partial h$ and

$(h_2-h_1)/(t_2-t_1)$. So the first order term on the right of (3) vanishes, on taking the mean, if the variations of $\partial\Theta/\partial h$, in time at a fixed point, are not correlated with the variations in $(h_2-h_1)/(t_2-t_1)$, at the same point and time.

In cumulus cloud eddies, the variations of velocity are caused by variations of the potential temperature $\Theta$, so that a correlation is almost certain to exist. On the contrary, when the eddies are due to dynamical instability, the correlation may be expected to vanish. In the latter case, it is the second order term of the right of (3) which becomes effective, so that

$$\overline{\frac{\delta\Theta}{\delta t}} = \overline{\frac{\partial^2\Theta}{\partial h^2}\cdot\frac{(h_2-h_1)^2}{2(t_2-t_1)}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

Now suppose further that either $\partial^2\Theta/\partial h^2$ has no variations at a fixed time and level, or else that its variations are not correlated with those of $(h_2-h_1)^2$. Then (9) simplifies to

$$\overline{\frac{\delta\Theta}{\delta t}} = \overline{\frac{\partial^2\Theta}{\partial h^2}}\cdot\frac{\overline{\{(h_2-h_1)^2\}}}{2(t_2-t_1)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

Thus

$$(h_2-h_1)^2/2(t_2-t_1) \text{ is the eddy-diffusivity K} \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

It is seen to be identical with that derived in Part IV. above, by considering the diffusion of a lamina, in which the density was distributed according to the law of error. It is a quantity easily measured.

Of course if $t_2-t_1$ were sufficiently small, say $\frac{1}{10}$ second, then it would be the first power of $h_2-h_1$, which would be proportional to $t_2-t_1$, instead of the square. This suggests that $t_2-t_1$ must be long compared with the fluctuations of the wind. On the other hand $t_2-t_1$ must be short compared with the period, of say 6 hours, over which the averages denoted by the bar are desired to be taken.

A similar argument can be applied to any other quantity which, like $\Theta$, does not change following the motion of the fluid, provided it has space-rates independent of the time-variations of velocity. Thus the mass-of-water-per-unit-mass-of-atmosphere may replace $\Theta$ in (10) with similar restrictions.

When we consider diffusion in three dimensions there may be six coefficients of diffusivity corresponding to the six components of stress.

## VI.—OSBORNE REYNOLDS' EDDY-STRESSES.

But we cannot, without further investigation, apply the preceding argument to the diffusion of horizontal velocity in a fixed azimuth.

Something might perhaps be deduced from the well-known theorem that, when $\rho$ is constant,

$$\frac{\mathrm{D}}{\mathrm{D}t}(\tfrac{1}{2}\rho v^2+\rho\psi+p) = \frac{\partial p}{\partial t} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

where $\psi$ is the gravity potential.

The eddy-viscosity can however be measured rigorously by smoke-puff observations made in such a manner as to fit in with OSBORNE REYNOLDS' theory of eddy-stresses.* This theory is remarkably free from assumptions which might limit its generality. It is to be found in LAMB'S 'Hydrodynamics,' IV. edn., Art. 369.

The equations of motion are three, such as

$$-\frac{\partial(\rho v_X)}{\partial t}=\frac{\partial p}{\partial x}+\frac{\partial}{\partial x}(\rho v_X v_X)+\frac{\partial}{\partial y}(\rho v_X v_Y)+\frac{\partial}{\partial h}(\rho v_X v_H)$$
$$+\rho\frac{\partial\psi}{\partial x}-2\omega\sin\phi v_Y-c\left(\tfrac{1}{3}\frac{\partial\,\mathrm{div}\,v}{\partial x}+\nabla^2 v_X\right)\quad .\quad .\quad .\quad .\quad (13)$$

where $c$ is the "molecular" or "ordinary" viscosity. Note that there is no need to assume $\rho$ to be independent of position. REYNOLDS assumed this, but for a reason that does not concern us. It will be necessary however to assume that $\rho'$, the variation of density at a fixed point, is so much smaller in comparison with $\bar{\rho}$ than is $v'$ in comparison with $\bar{v}$, that we may put $\rho'=0$. This being so, we find on taking the mean that (13) becomes

$$-\left\{\frac{\partial(\rho v_X)}{\partial t}\right\}=\frac{\partial\bar{p}}{\partial x}+\rho\frac{\partial\psi}{\partial x}-2\omega\sin\phi\bar{v}_Y-c\left(\tfrac{1}{3}\frac{\partial\,\mathrm{div}\,\bar{v}}{\partial x}+\nabla^2\bar{v}_X\right)$$
$$+\frac{\partial}{\partial x}(\rho\bar{v}_X\,.\,\bar{v}_X+\rho\overline{v'_X\,.\,v'_X})+\frac{\partial}{\partial y}(\rho\bar{v}_X\,.\,\bar{v}_Y+\rho\overline{v'_X\,.\,v'_Y})$$
$$+\frac{\partial}{\partial h}(\rho\bar{v}_X\,.\,\bar{v}_H+\rho\overline{v'_X v'_H})\quad .\quad .\quad .\quad .\quad .\quad .\quad .\quad .\quad .\quad (14)$$

The left side of (14) is the difference between $\rho v_X$ at the beginning and at the end of the period through which the average is taken, divided by the period; and that is what we want. The right side of (14) is of exactly the same form in the mean quantities $\bar{p}$, $\bar{v}_X$, $\bar{v}_Y$, $\bar{v}_H$ as (13) was in the corresponding instantaneous quantities $p$, $v_X$, $v_Y$, $v_H$; except that there is added a force per unit volume in the $x$ direction equal to minus

$$\frac{\partial}{\partial x}(\rho\overline{v'_X v'_X})+\frac{\partial}{\partial y}(\rho\overline{v'_X v'_Y})+\frac{\partial}{\partial h}(\rho\overline{v'_X v'_H}).\quad .\quad .\quad .\quad .\quad .\quad (15)$$

On working out the corresponding equations for the $y$ and $h$ components, it is seen that this additional force per unit volume is just that which would be given by the following systems of stresses

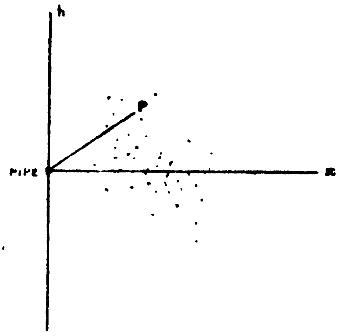
$$\left.\begin{array}{lll}\widehat{xx}=-\rho\overline{v'_X v'_X}\,; & \widehat{yy}=-\rho\overline{v'_Y v'_Y}\,; & \widehat{hh}=-\rho\overline{v'_H v'_H}\\ \widehat{xy}=-\rho\overline{v'_X v'_Y}\,; & \widehat{yh}=-\rho\overline{v'_Y v'_H}\,; & \widehat{hx}=-\rho\overline{v'_H v'_X}\end{array}\right\}\quad .\quad .\quad .\quad (16)$$

---

* Major G. I. TAYLOR tells me that he attempted to measure $\widehat{xh}$ with a balloon on an elastic tether in 1914.

when any symbol such as $\widehat{xy}$ is the force in the $x$-direction per unit area of a plane normal to the $y$ axis. Tractions are reckoned positive, as usual. The above is taken from OSBORNE REYNOLDS' theory, adapted and slightly generalized to suit our needs for a rotating atmosphere, having density diminishing with height and a molecular viscosity which is not neglected.

One may form a clear mental picture of these eddy-stresses by imagining the scattering of smoke puffs. Let a puff emerge from a pipe at the origin of the co-ordinates. After a short interval of time $\tau$, let the puff appear at the point P on the diagram, as seen by an observer at a distant point on the $y$-axis. Now let the observation be repeated for a large number of puffs in succession, the time $\tau$ being kept the same for each. We thus obtain a diagram, with a large number of points on it, showing the scattering of the puffs after $\tau$. Then the eddy-stresses are simply related to the correlations and standard deviations of this scatter-diagram—under certain conditions. For let X, Y, H now mean the co-ordinates of any one of the dots on the diagram reckoned from the source of smoke.



Then the velocities of the corresponding puff were

$$v_X = \frac{X}{\tau}; \qquad v_H = \frac{H}{\tau}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (17)$$

provided the time $\tau$ was so short that, during it, the velocity may be regarded as uniform and in a straight line. . . . . . . . . . . . . . . . . . . . (18)

Again, the velocity of the puff will be equal to that of the air which it has replaced provided the puff is at the same temperature as the air, and provided that the pipe points parallel to the Y-axis so that the impulse with which the puff leaves the pipe does not show in the projection on the plane XOH.

Let us suppose that a number of puffs, $n$ in all, are observed. In order to correspond with the time-mean taken over 6 hours, which was used in deriving the eddy-stresses from the equations of motion, these $n$ puffs should be spread uniformly over a similar interval.

From the scatter diagram we can compute first the mean velocities. For the mean velocities are

$$\bar{v}_X = \frac{\bar{X}}{\tau} = \frac{1}{\tau \,.\, n}\Sigma X; \qquad \bar{v}_Y = \bar{Y}/\tau = \frac{\Sigma Y}{\tau \,.\, n}, \quad . \quad . \quad . \quad . \quad (19)$$

where $\Sigma$ has the meaning :—take the sum of what follows it, for $n$ puffs.

Then the normal eddy-stresses, $\overset{\frown}{xx}$ and $\overset{\frown}{hh}$, are found thus

$$\overset{\frown}{xx} = -\rho\overline{v'_X v'_X} = -\rho\,\overline{(v_X-\bar{v}_X)(v_X-\bar{v}_X)}$$

$$= -\frac{\rho}{\tau^2 n}\Sigma(X-\bar{X})(X-\bar{X}). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

So

$$\overset{\frown}{xx} = -\rho\sigma_X^2/\tau^2, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (21)$$

where $\sigma_X$ is the "standard deviation" of the dots in the $x$-direction. If $\sigma_H$ is the corresponding quantity vertically

$$\overset{\frown}{hh} = -\rho\sigma_H^2\tau^{-2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (22)$$

So the direct eddy-stress in the direction of the wind is intimately related to the gustiness shown by a tube-anemometer.

The shearing eddy-stress

$$\overset{\frown}{xh} = -\rho\overline{v'_X v'_H} = -\frac{\rho}{\tau^2}\frac{1}{n}\Sigma(X-\bar{X})(H-\bar{H}). \quad . \quad . \quad . \quad . \quad . \quad . \quad (23)$$

So

$$\overset{\frown}{xh} = -\rho\tau^{-2}r_{XH}\sigma_X\sigma_H, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (24)$$

where $r_{XH}$ is the correlation between the co-ordinates X and H of the dots.

By projecting the puffs on the other two co-ordinate planes we should be able to measure similarly the remaining components of eddy-stress.

To find the eddy-viscosity we must compare the sheering eddy-stress $\overset{\frown}{xh}$ with $\left(\frac{\partial\bar{v}_X}{\partial h}+\frac{\partial\bar{v}_H}{\partial x}\right)$, which is the rate of shearing strain in the *mean* motion. Usually $\partial\bar{v}_H/\partial x$ is negligible, so that the rate of shearing is $\partial\bar{v}_X/\partial h$, a quantity which can easily be observed. At first sight one might think that $\partial\bar{v}_X/\partial h$ was simply related to the slope of the regression line in the scatter diagram; but on examination this proves not to be the case. The slope of the regression line is independent of $\tau$, because (18) is satisfied for all permissible intervals of time.

It should be noted that no shearing stress such as $-\rho\overline{v'_X v'_H}$ can exceed, in absolute value, the geometric mean of the corresponding pair of direct stresses $-\rho\overline{v'_X v'_X}$, $-\rho\overline{v'_H v'_H}$ for the same reason that a correlation coefficient cannot exceed unity.

The probable errors of eddy-stresses, determined from the scattering of particles moving with the air, may be taken to be as follows

$$\text{Probable error of } \overset{\frown}{xx} = 0{\cdot}674\,.\,\overset{\frown}{xx}\sqrt{\frac{2}{N}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

$$\text{Probable error of } \overset{\frown}{xh} = 0{\cdot}674\sqrt{\frac{\overset{\frown}{xh}^2+\overset{\frown}{xx}\,.\,\overset{\frown}{hh}}{N}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)$$

These are based on the assumption that the number of particles is not less than say 20, and that the scatter is "normal," so that the logarithm of the density of particles would be a quadratic function of $x$, $y$, $h$ if a very large number of particles were observed. Reference may be made to the fundamental papers on probable errors: FILON and PEARSON, 'Phil. Trans.,' A, vol. 191; W. F. SHEPPARD 'Phil Trans.,' A, vol. 192. Equation (26) follows from the probable error of the quantity called by K. PEARSON the "product moment coefficient *taken about the mean.*" Equation (25) follows from (26) on putting $x = h$, or may be deduced independently from the probable error of a standard deviation taken about the mean.

### *Corrections for the Motion of the Parachute Relative to the Air.*

When observing eddy-stresses by the aid of the parachutes of plant seeds it is desirable to allow for the velocity of the parachute in still air. For large specimens of the parachute of *Taraxacum officinale*, after cutting off the seed, the velocity in still air was found to have a mean of 12 cm. sec.$^{-1}$ with a standard deviation of 2 cm. sec.$^{-1}$. It may be shown from the equation of motion that this limiting velocity is acquired in a negligibly short interval of time. Thus call the limiting velocity $c$ downwards, the instantaneous velocity downwards $u$. Then if the friction is proportional to the velocity

$$g\,(\text{mass}) = c \times \mathrm{F}, \qquad (27)$$

where F is a constant. But, when accelerating,

$$(\text{mass})\,.\,\frac{du}{dt} = g\,(\text{mass}) - u\mathrm{F}. \qquad (28)$$

Eliminate the mass between these two equations and there results

$$\frac{c}{g}\,\frac{d\,(u-c)}{dt} = -\,(u-c). \qquad (29)$$

So that the discrepancy between the actual velocity $u$ and the terminal velocity $c$ sinks to $e^{-1}$ of itself in a time equal to $c/g$, which for the taraxacum parachute having $c = 12$ cm. sec.$^{-1}$ would be only a hundredth of a second.

Less negligible is the variation of the velocity $c$ from one parachute to another. What we actually observe is not $v_{\mathrm{H}}$ the upward velocity of the air, but $v_{\mathrm{H}} - c$. Now write $\bar{c}$ for the mean velocity of the parachute in still air, and $c'$ for the deviation from the mean. Then in finding the direct stress $\widehat{hh}$ we must perforce work out first the "raw" moment $\overline{\{(v_{\mathrm{H}}-c)^2\}} = \overline{[\{(\bar{v}_{\mathrm{H}}-\bar{c})+(v'_{\mathrm{H}}-c')\}^2]}$. On expanding and remembering that a bar put over the product of a dashed and a barred symbol, causes the result to vanish, and also that $\overline{v'_{\mathrm{H}}-c'} = 0$, it is found that

$$\overline{\{(v_{\mathrm{H}}-c)^2\}} = (\bar{v}_{\mathrm{H}}-\bar{c})^2 + \overline{v'_{\mathrm{H}}v'_{\mathrm{H}}} + \overline{c'c'}$$

it follows that the corrected stress is given by

$$\widehat{hh} = -\rho\overline{v'_{\mathrm{H}}v'_{\mathrm{H}}} = -\rho\left\{\overline{(v_{\mathrm{H}}-c)(v_{\mathrm{H}}-c)}-(\bar{v}_{\mathrm{H}}-\bar{c})^2-\overline{c'c'}\right\} \quad . \quad . \quad . \quad . \quad (30)$$

which is the actual formula used in working up the observations. Here $\overline{c'c'}$ is the square of the standard deviation of the velocity of the parachutes in still air.

Next we require the product moments in order to find the shearing stresses $\widehat{xh}$ and $\widehat{yh}$. The raw moment $\overline{v_{\mathrm{X}}(v_{\mathrm{H}}-c)}$ is found on expanding to be equal to $\bar{v}_{\mathrm{X}}\overline{(v_{\mathrm{H}}-c)}+\overline{v'_{\mathrm{X}}v'_{\mathrm{H}}}$.

So that the corrected value for the stress is

$$\widehat{xh} = -\rho\overline{v'_{\mathrm{H}}v'_{\mathrm{X}}} = -\rho\left\{\overline{v_{\mathrm{X}}(v_{\mathrm{H}}-c)}-\bar{v}_{\mathrm{X}}\overline{(v_{\mathrm{H}}-c)}\right\}, \quad . \quad . \quad . \quad . \quad . \quad (31)$$

and as $\overline{c'c'}$ does not appear, the scatter of the velocities of the parachutes in still air does not make a correction necessary for the shearing stresses.

## VII. Summary of Theory of Scattering of Particles of Air.

The conclusion we have reached is the following. For any sort of eddy, whether due to "dynamical instability," or to the rising of heated air in cumuli, the eddy-stresses are best measured by equations (22), (24) and the like, because the theory from which they are derived is very general; and the eddy-viscosity is best measured as the ratio of the shearing eddy-stress to the rate of mean shearing strain. It is conceivable that $\widehat{xh}$ found from (24) might turn out to be zero. In that case it would be necessary to investigate effects of higher order. This might possibly be done by developing, for the quantity $(\frac{1}{2}\rho v^2+\rho\psi+p)$ in equation (12) an analysis similar to that of (1) to (11) for potential temperature. The diffusivity for potential temperature, on the other hand, should be measured differently according as the eddies are produced by variations of potential temperature or not. Thus for cumulus eddies we should take the mean of (3), retain the linear term on its right-hand side and neglect the quadratic one. Then $t_2-t_1$ must be small, so that

$$\overline{\frac{\eth\Theta}{\eth t}} = \overline{\frac{\partial\Theta}{\partial h}\cdot\frac{(h_2-h_1)}{t_2-t_1}} = \overline{\left(\frac{\partial\Theta}{\partial h}\right)'v'_{\mathrm{H}}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (31\mathrm{A})$$

The diffusivity is measured as the right side of this equation divided by $\overline{\partial^2\Theta/\partial h^2}$. Thus

$$\mathrm{K} = \frac{\overline{\left(\frac{\partial\Theta}{\partial h}\right)'v'_{\mathrm{H}}}}{\overline{\partial^2\Theta/\partial h^2}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (32)$$

But for eddies due to dynamical instability, neglect the linear term in (3) and measure the diffusivity as

$$\mathrm{K} = \frac{\overline{(h_2-h_1)^2}}{2(t_2-t_1)}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (33)$$

where $(t_2-t_1)$ must not be too small. It will be interesting to see whether eddy-diffusivity is found to be equal to eddy-viscosity divided by density.

Let us now see how these two modes of disposal of smoke fit together—the one, at short intervals of time, governed by the eddy-stresses, the other at long intervals governed by the diffusivity. Let us draw the trail of smoke as it would appear to a distant observer looking, say, horizontally. The scatter may be represented by drawing two lines through the points where the smoke-density has its standard deviation in height above or below its mean. Let T be the time since the smoke emerged from its source. Suppose that the stress $\widehat{hh}$, the diffusivity K and the mean velocity $\bar{v}_{\mathrm{x}}$ are all constant along the path of the smoke. The time T taken to travel a horizontal distance $x$ measured down the trail from the source is $x/\bar{v}_{\mathrm{x}}$. Near the origin $\mathrm{T} = \tau$ in equation (22), and so the standard deviation in height is

$$\pm \frac{x}{\bar{v}_{\mathrm{x}}} \sqrt{-\widehat{hh}/\rho} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (34)$$

representing a pair of straight lines intersecting at the source. Further down the trail $\mathrm{T} = t_2 - t_1$ in equation (11), and the standard deviation in height is

$$\pm \sqrt{x} \cdot \sqrt{\frac{2\mathrm{K}}{\bar{v}_{\mathrm{x}}}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (35)$$

representing a parabola with horizontal axis and its apex at the source. Thus according to the theory, the smoke may be said to be contained within a paraboloid which has had its blunt end sharpened into a cone. The preceding theory gives us no clue as to the manner of transition from the cone to the paraboloid.

When we can observe a sufficient length of *the path traced out in space by a single small portion of air* the eddy-stresses and the eddy-diffusivity may be deduced from the irregularities in the motion. With this object I have observed the motion of anti-aircraft shell bursts, and of portions of cloud, by means of an Abney level or a pocket sextant. With better instruments this method might yield a good deal of information about eddies at heights such as 2 to 5 km. One principal difficulty is that the shell-burst fades away after about 5 minutes, before a sufficient length of path has been observed to give the diffusivity.

If the path is sufficiently high and long, the hills, trees and houses on the earth may be regarded as blending into a "roughness." Suppose this roughness to be uniform. Then if we had been causing smoke to issue in puffs from a fixed pipe, we should presumably have obtained the same scatter diagram for the puffs, within the limits of probable error, at whatever point of the path we had placed the pipe, or at whatever time we had begun to observe, within limits. If this is so, we may form the scatter diagram by taking its origin at every point in succession of the trajectory of the single particle. For instance, Captain CAVE in his book on "The Structure of the Atmosphere," gives several diagrams of the irregularities of height of a balloon observed by two theodolites, the uniform vertical motion of the balloon relative to the air having been eliminated. From his figure 30, of an ascent on February 19, 1909, the following has been deduced, by taking the origin of scattering at every available minute mark, and working out the standard deviation, in height about the mean height, after a time $\alpha$.

TABLE III.

| $\alpha$, secs. | $1 \times 60$ | $3 \times 60$ | $5 \times 60$ | $7 \times 60$ | $9 \times 60$ |
|---|---|---|---|---|---|
| $\frac{\sigma_H^2}{2\alpha} = 10^4 \times$ cm.$^2$ sec.$^{-1}$ | 0·93 | 0·82 | 0·55 | 0·55 | 0·33 |
| $\rho\frac{\sigma_H^2}{\alpha^2}$ dyne cm.$^{-2}$ | 0·345 | 0·100 | 0·041 | 0·029 | 0·014 |
| Number of points | 14 | 12 | 10 | 8 | 6 |

Now $\sigma_H^2/2\alpha$ would be the diffusivity if $\alpha$ were a "long" time; and the criterion of sufficiency in length is that $\sigma_H^2/2\alpha$ should not vary with $\alpha$. The small number of points makes the probable errors large, so that the decrease of $\sigma_H^2/2\alpha$ between $\alpha = 7$ mins. and $\alpha = 9$ mins. is not significant. It looks as though the diffusivity K were here of the order of $0{\cdot}5 \times 10^4$ cm.$^2$ sec.$^{-1}$.

Again $-\rho\sigma_H^2/\alpha^2$ would be the eddy-stress, $\widehat{hh}$, if $\alpha$ were so small that further decrease made no further change in the quantity. This stage is not reached at $\alpha = 1$ minute. All that we can say is that the eddy-stress is probably greater than 0·345 dynes cm.$^{-2}$.

The mean height of this observation is 1 km. above ground, and the mean velocity 1270 cm. sec.$^{-1}$.

The photograph of a smoke trail in fig. 2 suggests that the eddying is partly random but also partly sinusoidal. Let us therefore see what would happen *if the path of the particle were an exact sine curve* without any random variations. Let the height $h$ of the particle be given by $h - \mathrm{B} = \mathrm{A} \sin qt$ where A, B and $q$ are constants. The increase in height in a time $\alpha$ would be

$$\mathrm{A}\left[\sin\left\{q\left(t+\frac{\alpha}{2}\right)\right\} - \sin\left\{q\left(t-\frac{\alpha}{2}\right)\right\}\right] = \mathrm{A} \,.\, 2 \cos qt \,.\, \sin\left(q\frac{\alpha}{2}\right)$$

by trigonometry. So that the standard deviation $\sigma_H$ after a time $\alpha$ would be given by

$$\sigma_H^2 = \frac{4\mathrm{A}^2\left(\sin\frac{q\alpha}{2}\right)^2}{\mathrm{L}} \int_{t=0}^{t=\mathrm{L}} (\cos qt)^2\, dt$$

where L is a very long time. The integral is equal to $\frac{1}{2}$L plus a negligible oscillatory part. Consequently $\sigma_H^2 = 2\mathrm{A}^2\left(\sin\frac{q\alpha}{2}\right)^2$. It follows that the stress $\widehat{hh}$, which is the limit of $-\rho\sigma_H^2/\alpha^2$ when $\alpha$ is small, comes to $-\rho q^2\mathrm{A}^2/2$; whereas the diffusivity K, which is the limit of $\sigma_H^2/2\alpha$ when $\alpha$ is long, comes to zero.

### *Comparison with* TAYLOR'S *expression for the Diffusivity and with* W. SCHMIDT'S *"Austausch."*

In TAYLOR'S remarkable investigation ('Phil. Trans.,' A, vol. 215) from which the present research took its stimulus, the diffusivity K is given, in the present notation, as the mean value of $v_H(h-h_0)$ over a large horizontal plane; and it is stated that $h-h_0$ is the height through which an eddy moves from the layer at which it was at the same temperature as its surroundings, to the layer with which it mixes. This definition of $h-h_0$ is puzzling, for it seems impossible to reconcile the supposed starting and stopping of the air, with the ceaseless motion which we observe in nature, except in the case of cumulus eddies. Happily we are now in a position to clear away the mystery. For it has been shown independently in the present paper that the diffusivity is given by

$$K = \overline{(h_2-h_1)^2/2\,(t_2-t_1)}$$

where $(t_2-t_1)$ is a time long compared to the fluctuations in the wind, and where the bar implies an average taken over a still much longer time. As $(t_2-t_1)$ is the same for all the quantities which are averaged, we may remove it from under the bar, writing

$$K\,(t_2-t_1) = \tfrac{1}{2}\,\overline{(h_2-h_1)^2}.$$

Differentiate this equation with respect to $t_2$,

$$K = \overline{(h_2-h_1)\,\frac{d}{dt_2}\,(h_2-h_1)} = \overline{(h_2-h_1)\,.\,v_H} \textit{ at } t_2,$$

thus K is expressed as the mean of the product of the rise in height during a long time into the vertical velocity *at the end* of that time. It may also be taken at the beginning. Comparing with TAYLOR'S form quoted above we see a strong resemblance, and we are led to suppose that TAYLOR'S theory makes two unnecessary and unnatural restrictions: (1) that the portion of air should start at the same temperature as its surroundings; (2) that the portion of air should finally mix with its surroundings. But that if these restrictions be removed, then another becomes necessary, namely that $(t_2-t_1)$ should be sufficiently long (several minutes). Whether the average be taken over a large horizontal plane, or over a very long time (6 hours), appears to be a matter of indifference.

The extent to which TAYLOR assumes viscosity to be independent of height in his *general* theory ('Phil. Trans.,' A, vol. 215, pp. 11 to 13) is this: he neglects the terms due to the initial eddying in his equation (6). That is a doubtful proceeding, unless the initial eddying is zero: but zero is independent of height.

The "Austausch" of W. SCHMIDT is defined by him (in 'Sitz. Akad. Wiss.,' Wien (1917), pp. 4 to 5) as

$$\frac{\Sigma\,(\text{element of mass crossing horizontal plane}) \times (\text{vertical displacement of element})}{(\text{whole area}) \times (\text{time of motion})}.$$

If we replace the summation by an integration over a large area S in the plane of $xy$, the element of mass crossing per unit time is $dx\,dy\,\rho v_{\text{H}}$, so that the "Austausch" becomes $\frac{1}{S}\iint \rho v_{\text{H}}\,(h-h_0)\,dx\,dy$ which is $\rho$ times TAYLOR's diffusivity as defined in 'Phil. Trans.,' A, vol. 215, p. 3. Thus we must suppose that W. SCHMIDT's definition of "Austausch" requires amplification concerning the interval of time and concerning the position in it of the velocity, just as TAYLOR's definition of K does.

## VIII. NUMERICAL VALUES DERIVED FROM THE SCATTERING OF PARTICLES.

Fig. 1 is a photograph* of the trail of paraffin vapour from an extinguished blast-lamp which projected the vapour in a direction at right angles to the wind. It shows a cone, with a blunt point due to the finite size of the source of smoke, passing smoothly into a form, which certainly diverges less rapidly than the initial cone, and which looks like a paraboloid. Opinion might differ slightly as to where to draw the lines corresponding to the standard deviation of smoke. In a "normal" distribution 0·68 of the whole number of particles lie between the two standard deviations. If the lines are placed as in the accompanying black and white drawing, then it follows, as the mean velocity of the smoke was 1·7 metres/sec., and the density of the air was $1{\cdot}21 \times 10^{-3}$ grm. cm.$^{-3}$, that

$$\text{stress}\ \widehat{hh} = -0{\cdot}73 \text{ dyne cm.}^{-2}\quad \text{diffusivity} = \text{K} = 240 \text{ cm.}^2 \text{ sec.}^{-1}$$

$$\text{turbulivity} = \xi = 340 \text{ grm.}^2 \text{ cm.}^{-2} \text{ sec.}^{-5}.$$

This photograph was taken in the evening, when the day-wind was diminishing. The source was 190 cm. above ground. Obstructions to windward only subtended an angle of 2°·1 at the source of smoke. The exposure lasted 60 seconds.

Fig. 2 was taken five minutes later in the same place, with an exposure of 85 seconds.

The velocity of the smoke had decreased to 1·3 metres per sec. The measurements yield

$$\widehat{hh} = -1{\cdot}2 \text{ dynes cm.}^{-2};\qquad \text{K} = 750 \text{ cm.}^2 \text{sec.}^{-1};\qquad \xi = 1050 \text{ grm.}^2 \text{cm.}^{-2} \text{sec.}^{-5}.$$

In this case the photograph shows a distinct neck between the cone and the paraboloid, at a distance from the source roughly 1·3 times its height above ground. This neck can also be recognized in some other photographs. Its presence signifies that the motion of the air was compounded of (i.) a random eddying, plus (ii.) a wave motion in which the particle of air executed a wave having a length, relative to a point fixed to the earth, roughly 2·6 times the height of the particle above ground.

---

* Taken at Benson.

### TABLE IV.—Observed Values of

The co-ordinate axes are taken to point: $0x$ horizontally with the mean wind at quantities are in C.G.S. units. Tractions are reckoned positive, so that $\widehat{xh}$ and $\widehat{yh}$

| Entropy gradient vertically upwards. | Surface. | Methods | Notes. | $h$ cm. above surface. | $\bar{v}_x$ $\frac{\text{cm.}}{\text{sec.}}$ | $\widehat{xx}$ dynes per square centimetre. |
|---|---|---|---|---|---|---|
| Positive sunrise . . | Tall grass . . . | Smoking wick . . . | I. | 30<br>170<br>300,000 | 10<br>60<br>200 | — |
| Small ? . . . . . | Oat field . . . | $NH_4Cl$ puffs . . . .<br>$NH_4Cl$ puffs . . . .<br>Dandelion down . . | II. | 40<br>100<br>50 | 20<br>40<br>— | —<br>—<br>— |
| Small ? . . . . . | Corn field . . . | Smoke puffs . . . . | III. | 120 | 100 | − 0·1 |
| Small ? . . . . . | Moor . . . . | Smoking wick . . . | IV. | 165 | 100 | — |
| ? . . . . . . | Moor, trees . . | Thistle down . . . | V. | 200 | 145. | − 2·4<br>± 0·5 |
| Positive ? . . . . | Flat field . . . | Paraffin vapour . . . | VI. Figs. 1 and 2 | 190 | 170<br>140 | —<br>— |
| 1919, Mar. 4<br>Between England and the Isle of Wight. | Sea . . . . . | Steamer's smoke . . | | 4,000 | 100 ? | — |
| 1919, Apr. 7, 18h. | Sea . . . . . | Steamer's smoke . . | | 10,000 | 200 ? | — |
| 1919, July 4, 11h. overcast | Sea . . . . . | Steamer's smoke . . | | 2,000 | 130 | — |
| 1919, July 5, 12h. | Sea . . . . . | Steamer's smoke . . | | 1,400 | 140 | — |
| | Ditcham . . . | Capt. CAVE's balloon . | p. 16 | 100,000 | 1270 | — |
| | Flat fields . . . | Phosphorus pentoxide | VI. Fig. 3 | 340 | 120 | — |

*Eddy-stresses* and of *Eddy-diffusivity.*

the level of observation, 0*y* horizontally to the left and 0*h* vertically upwards. All are positive when the air above drags the air below in senses of *x* and *y* increasing.

| $\widehat{yy}$ dynes per square centimetre. | $\widehat{hh}$ dynes per square centimetre. | $\widehat{xy}$ dynes per square centimetre. | $\widehat{yh}$ dynes per square centimetre. | $\widehat{hx}$ dynes per square centimetre. | $\frac{\partial \bar{v}_r}{\partial h}$ sec.$^{-1}$. | Conductivity $= \rho K$ grm. cm.$^{-1}$ sec.$^{-1}$. | Diffusivity. K cm.$^2$ sec.$^{-1}$. | Turbulivity. $\xi = g^2\rho^2K$ grm.$^2$ cm.$^{-2}$ sec.$^{-5}$. |
|---|---|---|---|---|---|---|---|---|
| — | — | | — | — | >0·5 | 0·006 | 5 | 7 |
| — | 0·004 | | | — | 0·3 | | | |
| — | 0·006 | | | — | | | | |
| — | — | | | | | 0·07 | 60 | 80 |
| — | −0·2 | | | <0·1 | <0·06 | | | |
| — | −0·04 | — | | — | — | 0·03 | 24 | 34 |
| −2·9<br>± 0·6 | −0·6<br>±0·12 | [+0·45]<br>±0·39 | −0·48<br>±0·20 | [−0·34]<br>±0·18 | | | | |
| — | −0·7 | — | — | — | — | 0·3 | 240 | 340 |
| — | 1·2 | | | | | 0·9 | 750 | 1,050 |
| — | — | — | — | | — | 12 | 10,000 | 14,000 |
| — | — | — | — | — | — | 0·25 | 200 | 300 |
| — | — | — | — | — | — | 6·8 | 5,500 | 8,000 |
| — | — | — | — | — | — | 5·9 | 4,800 | 7,000 |
| — | < −0·34 | — | — | — | — | 6 | 5,000 | 6,000 |
| — | — | — | — | — | — | 0·16 | 130 | 200 |

TABLE IV.—Observed Values of

| Entropy gradient vertically upwards. | Surface. | Methods. | Notes. | $h$ cm. above surface. | $\bar{v}_x$ $\frac{\text{cm.}}{\text{sec.}}$. | $\widehat{xx}$ dynes per square centimetre. |
|---|---|---|---|---|---|---|
| 1919, Oct., 21d. 12h., BENSON | Obstructions up wind $\frac{1}{10}$ radian | Paraffin vapour . . . | | 160 | 250 | — |
| 1919, Oct., 29d. 15h., BENSON; raining | Obstructions up wind 0·045 radian | Smoke . . . . . | | 190 | 330 | — |
| Small ? . . . . . | Moor, trees . . | Burning rubbish . . | VII. | 900 | 300 | — |
| Small ? . . . . . | Fields, trees . . | Factory chimney . . | VIII. | 3,000 | 600 | — |
| Negative ? . . . . | Moor, low hills . | Cloud . . . . . . | IX. | 150,000<br>2,000 | 1000<br>200 | —<br>— |
| Small ? . . . . . | Wooded hills . . | Large fire . . . . | X. | 25,000<br>37,000 | —<br>1600 | — |
| | | Shell-puffs . . . . | XI. | 300,000 | 500 | — |
| Positive ? . . . . | Spurn head . . | Dines anemometer . | XII. | — | 1200 | − 80 |
| Positive ? . . . . | Open sea . . . | Steamer's smoke . . | XIII. | 1,500 | — | — |
| Negative ? . . . . | Open sea . . . | Steamer's smoke . . | XIV. | 5,000 | 200 | — |
| ? . . . . . . . | Open sea . . . | Alto-stratus . . . . | XV. | 500,000 | — | — |

*Notes to Table of Eddy-Stresses and Diffusivities.*

I. 1917, June, 16d. 4h. 8m. L.A.T., hilltop near Ancemont, France. Standing hay composed of a species of *Festuca* (identified by my friend, Mr. SAM PIM). It grew fairly densely to 30 cm. from the ground and tall seed stems rose to 70 cm.

II. 1917, July, 16d. 19½h. L.A.T., Maffrecourt, France. Green oats 70 cm. high. No trees near. About sunset. Overcast with stratus. Observers: DAVID LONG and L. F. RICHARDSON.

III. 1917, June, 29d. 19h. 30m. L.A.T., Viel Dampierre, France. A field of corn 60 cm. high. Clouds (stratus) motionless. Observers: F. H. WEATHERALL, G. HUTCHINSON, L.F.R.

*Eddy-stresses* and of *Eddy-diffusivity* (continued).

| $\widehat{yy}$ dynes per square centimetre. | $\widehat{hh}$ dynes per square centimetre. | $\widehat{xy}$ dynes per square centimetre. | $\widehat{yh}$ dynes per square centimetre. | $\widehat{hx}$ dynes per square centimetre. | $\frac{\partial \bar{v}_x}{\partial h}$ sec.$^{-1}$ | Conductivity. $= \rho K$ grm. cm.$^{-1}$ sec.$^{-1}$. | Diffusivity. K cm.$^2$ sec.$^{-1}$. | Turbulivity. $\xi = g^2\rho^2 K$ grm.$^2$ cm.$^{-2}$ sec.$^{-5}$. |
|---|---|---|---|---|---|---|---|---|
| — | — | — | — | — | — | 2·5 | 2,000 | 3,000 |
| — | — | — | — | — | — | 9·3 | 7,500 | 11,000 |
| — | — | — | — | — | — | 6 | 5,000 | 7,000 |
| — | — | — | — | — | — | 12 | 10,000 | 14,000 |
| —<br>— | -5<br>— | | | | | | | |
| — | — | — | — | — | | 120 | $10^5$ | $1·4 \times 10^5$ |
| — | — | — | — | — | — | $< 0·9 \times 10^3$ | $< 10^6$ | $< 0·8 \times 10^6$ |
| -110 | | | | | | | | |
| — | — | — | — | — | — | 0·12 | 100 | 140 |
| — | — | — | | — | — | 12 | 10,000 | 14,000 |
| — | -1 | | | | | | | |

IV. 1917, June, 25d. 20h. 5m. L.A.T., Joinville, France. Moor with herbage dense to 10 cm. and rising thinly to 50 cm.

V. 1917, October, 4d. 8½h. G.M.T., Massiges, France. Flat moor with grass to 10 cm. and stems rising to 30 cm. Trees up-wind subtending an angle of 10 degrees. Overcast with strato nimbus, of which velocity/height = 0·025 sec.$^{-1}$. Temperature 0°·15 C. Observers: OLAF STAPLEDON and L. F. RICHARDSON. Eddies partly due to observer.

VI. See photographs and description in this paper.

VII. 1917, July, 18d. 7¼h. L.A.T., Maffrecourt, France. Moor, with small trees, 5m. high and houses. Overcast.

VIII. 1917, August, 2d. 17h. L.A.T., South England. Overcast.

IX. 1917, July, 17d. 19h. 50m. L.A.T., East Champagne, France. Nine observations of a cloud at intervals of 1 minute by an Abney level. Angular elevation about 9 degrees.

X. 1917, August, 8d. 20h. L.A.T., Argonne, France. Large fire of petrol and wood. Smoke observed from distance of 10 km. with a sextant. If smoke were not hot, height of its upper edge above ground would have given $K = 4 \cdot 6 \times 10^5$ cm.$^2$ sec.$^{-1}$, an over estimate. Irregularities in upper edge gave $K = 10^4$, probably an under estimate. Mean K of order of $10^5$. Overcast.

XI 1918, April, 12d. 14h. 5m. L.A.T., France. Two anti-aircraft shell puffs at a mean elevation of 21 degrees above the horizon and 4 degrees apart in a vertical plane, were brought into coincidence in the field of view of a sextant. In 120 seconds their separation of 4 degrees did not vary visibly, certainly not by 4 minutes of arc. Their apparent motion was horizontal at $\frac{1}{1800}$ radian per sec. Height *assumed* 3200 metres—a likely value.

XII. 1914, November, 16d. 3h. to 9h. G.M.T. Taken from the re-production of the Dines anemogram on p. 81 of the 'Observer's Handbook,' Meteorological Office, London, 1917 edition.

XIII. 1917, July, 26d. 21h. L.A.T., the English Channel.

XIV. 1917, July, 26d. 16h. L.A.T., the English Channel, off Havre.

XV. 1917, July, 26d. 15h. L.A.T., the English Channel, off Havre. Observations of cloud at intervals of 1 minute with pocket sextant. Angular elevation about 4 degrees.

The photograph (2) also shows that the smoke spreads more rapidly upwards than downwards, indicating that the stress $\widehat{hh}$ and the turbulivity $\xi$ both increase with height.

Fig. 3 shows another case of low eddy-conductivity occurring at sunset. The smoke here is from burning hydrogen phosphides; it is warm and rises slightly. To the eye the smoke appeared as a narrow wavy ribbon moving with a mean velocity of 1·2 metres per second. The broader smooth band shown in the photograph is due to the exposure of 75 seconds, made long in order to get an average effect. The source of smoke is a bottle 3·4 metres above ground and just within the picture. The bamboos are 5 metres apart. The air density was 0·00126 c.g.s. The eddy-diffusivity works out to about 130 c.g.s. units, the eddy-conductivity to 0·16 c.g.s., the turbulivity to 200 c.g.s. The sky was cloudless. Obstructions to windward rose above the horizon to an angle of only $\frac{1}{50}$ radian. The photograph was taken in latitude 51° 37′ N., longitude 4m. 24s. west, at 1919, Sept., 29d. 17h. 58m. G.M.T.

Above is a table of observations. It is noticeable that when two of the direct stresses $\widehat{xx}$, $\widehat{yy}$, $\widehat{hh}$ have been measured at the same time and place, they have been found to be not very unequal. G. I. TAYLOR has published some observations which show the same thing. It is as though there were a kind of equipartition of energy between the three components of the eddying motion. A very marked increase in both direct stress and diffusivity takes place either with velocity or with height. A rapid increase of viscosity with height in the first 200 metres has also been deduced by W. SCHMIDT from wind observations made by HELLMANN over a piece of flat land. ('Sitz. Akad. Wiss.,' Wien, 1917, Heft 6, p. 17).

The six observations of steamers' smoke are put forward only as upper limits to the turbulence appropriate to the bare sea, for the steamer itself probably makes a considerable eddy.

It would be desirable to classify the observed eddy-conductivity as a function of four independent variables; namely, the height, the vertical gradient of entropy, the vertical gradient of velocity and the character of the surface. Vertical gradient of velocity is suggested as an independent because it measures the only rate-of-mean strain which attains a noticeable value in the free atmosphere, and because **Osborne Reynolds**[*] has shown that the energy of the eddy motion comes from the work done by the eddy-stresses upon the corresponding rates of mean strain. The observations here presented are much too scanty for such a classification, but to render the relation to height visible, the effect of velocity has been removed, in one sense, by dividing each value of the eddy-conductivity by the velocity at that level. The

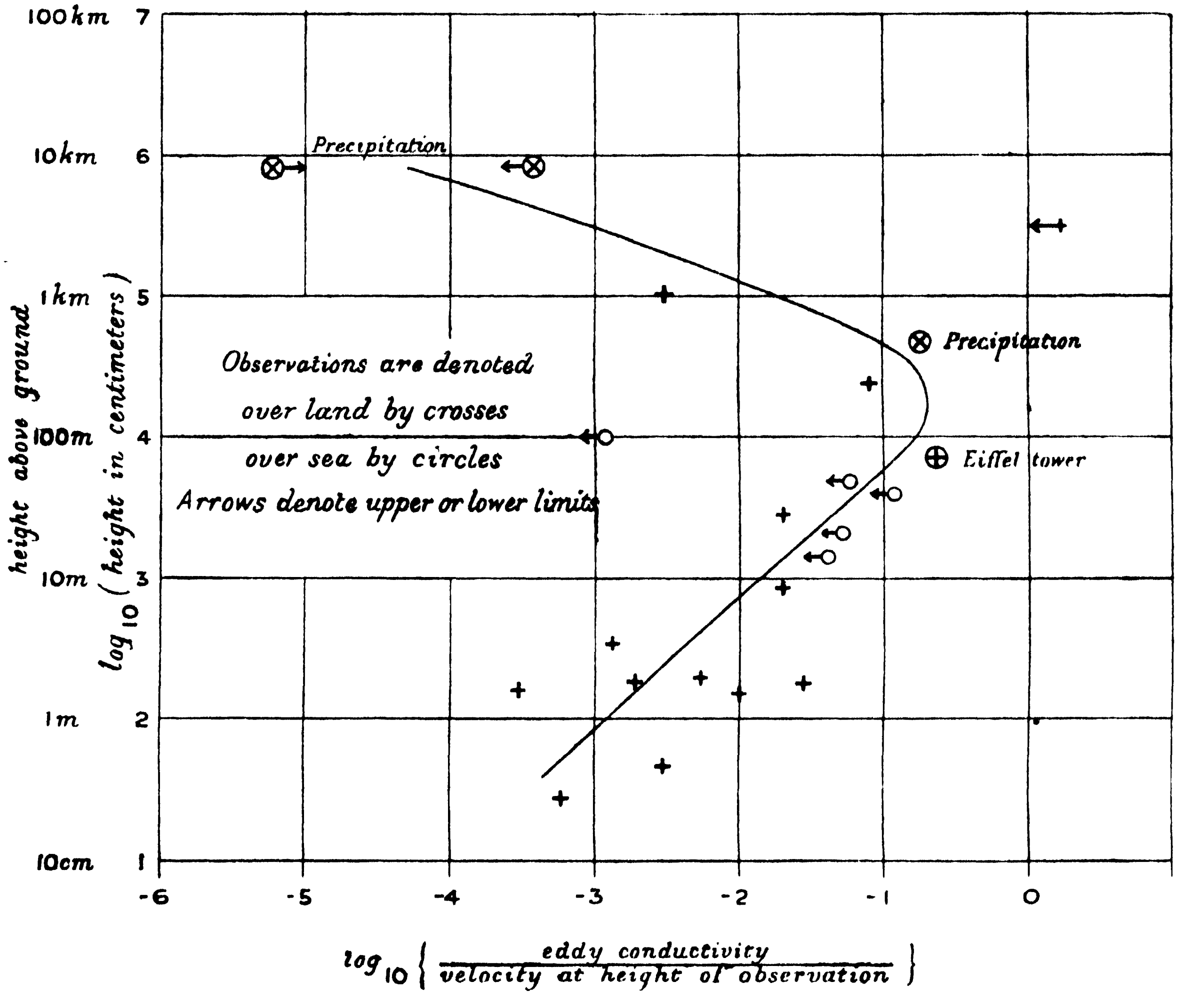


Fig. 4.

---

[*] **Lamb**, 'Hydrodynamics,' IV. edition, § 369, equation (21).

justification for this procedure is that TAYLOR has given reasons[*] for supposing that the viscosity, and therefore also the conductivity, is proportional to the velocity. For comparison with the present observations, the diagram shows TAYLOR'S mean value of the diffusivity at the Eiffel Tower,[†] and also some general means[‡] deduced from precipitation by the writer.

In order to compare them it has been necessary to assume some corresponding velocities; for which purpose I have taken 540 cm. sec.$^{-1}$ at the mean height of the Eiffel Tower, 700 cm. sec.$^{-1}$ as a world-mean at 500 metres and 1000 cm. sec.$^{-1}$ for the same at 8500 metres. These are based on information given in HANN'S 'Meteorology.' The conversion formulæ between eddy-conductivity, diffusivity and $\xi$ have been given in Section I. In order to compress into a diagram the large ranges of height and conductivity, logarithms have been plotted. A smooth curve is drawn through the clustered observations over land. It shows a maximum between the heights of 100 and 1000 metres, and a marked falling off above and below. Not only $c/v$ but also $c$ the conductivity has a maximum here. TAYLOR'S first observations related to heights near this maximum and so he naturally came to the conclusion that there was no marked variation with height.

## IX. CUMULUS EDDIES IN CALM WEATHER.

The familiar sequence, which can be observed in many places, is here illustrated by the mean of some selected days in latitude 49° in France, on the bare grass moors to the west of the forest of Argonne, in the month of May. The sun rose at 4h. 20m. local apparent time, but could not be seen for mist. By 6h. the disk of the sun became visible. At 7½h. the mist was rising in large pieces, leaving a brilliant blue sky. At 9h. the first cumuli appeared over the forest. About half-an-hour later they appeared over the grass land also. By noon the cumuli covered $\frac{4}{10}$ of the sky. By 16h. the cumuli had begun to spread out horizontally, and by 19h. they had vanished, leaving the sky clear again.

Now here we have a collection of eddies in which the rising parts, represented by the cumuli, visibly move to a level where they remain by mixing with their surroundings. So we should be able to calculate the diffusivity K by the direct application of the formula given by G. I. TAYLOR ('Phil. Trans.,' A, vol. 215, p. 3)

$$K = \frac{1}{A}\iint_A v_H (h-h')\,dx\,dy, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where $h-h'$ is the height through which the air has moved before mixing, $v_H$ is its vertical velocity, and A is a large horizontal area. Only, as TAYLOR'S formula assumes that $\rho$ and K are independent of height, it may be as well to remove these restrictions. It is then found that $\xi$, as defined by equation (1), of Part I., is given by

---

[*] G. I. TAYLOR, 'Roy. Soc. Proc.,' A, vol. 92, pp. 196–199.

[†] G. I. TAYLOR, 'Roy. Soc. Proc.,' A, vol. 94 (1917), p. 141.

[‡] L. F. RICHARDSON, 'Roy. Soc. Proc.,' A, vol. 96 (1919), p. 18.

$$\xi = \frac{g}{\mathrm{A}} \iint_{\mathrm{A}} m_{\mathrm{H}} (p'-p)\, dx\, dy, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (2)$$

where $p'$ is the pressure at the initial level, $m_{\mathrm{H}}$ the vertical momentum per volume.

Now let us insert numerical values. The surface air was seen to begin to move at or a little before $7\frac{1}{2}$h. The cumuli appeared to cease rising before 16h. and the height of their tops is known to average about 2 km. (*vide* HANN's 'Meteorology,' IIIrd edn., p. 280). Now if we suppose, as seems reasonable, that the top of the cumulus is formed from the damp air which was initially close to the ground, then the displacement, measured by pressure, is about 2 decibars, so that $(p'-p) = 2 \times 10^5$ dyne cm.$^{-2}$. The vertical velocity is 2 km. in 8·5 hours, that is 6·5 cm. sec.$^{-1}$. So the momentum per volume $= m \;\; = 7 \times 10^{-3}$ grm. cm.$^{-2}$ sec.$^{-1}$ in the rising current. The rising current covered 0·4 of the sky, so that averaging over the area A, as is done in (2), is equivalent to taking 0·4 of $m_{\mathrm{H}} (p'-p)$ for the rising current. But the invisible descending currents contribute an equal amount to the integral. So

$$\begin{aligned} \xi &= g \times 0{\cdot}8 \times 2 \times 10^6 \times 7 \times 10^{-3} \\ &= 11 \times 10^5 \text{ grm.}^2 \text{ cm.}^{-2} \text{ sec.}^{-5}. \end{aligned}$$

This figure is about ten times greater than measures of $\xi$ at a height of a few hundred meters, deduced by various authors. If the air which forms the top of the cumulus had really started from a height of 1 km. instead of from the ground, as we have supposed, then the numerical value of $\xi$ would have to be divided by four.

Reasons have already been given (Part VII.) for supposing that $\xi$ derived in this way from cumulus clouds is a measure of frictional effects, but not of the diffusion of entropy, because the linear term in (Part V., 3) does not vanish on taking the mean, owing to the fact that the eddies are produced by variations of entropy. To put it in another way: In G. I. TAYLOR's deduction of formula (1) the vertical gradient of the diffusing quantity is treated as not correlated with the vertical velocity. When we are dealing with cumulus clouds that assumption is probably justified if the diffusing quantity is horizontal velocity, but not if it is potential temperature.

To find $\xi$ in the sense of diffusivity for potential temperature we should have to employ formula 32 of Part VII., namely

$$\mathrm{K} = \frac{\overline{\left(\dfrac{\partial \Theta}{\partial h}\right)' v'_{\mathrm{H}}}}{\overline{\partial^2 \Theta / \partial h^2}}.$$

For insertion in this we require lapse rates in cumulus clouds and in the clear air between them. Such have recently been obtained by airmen.

## X. SUMMARY.

Part I. deals with notation. Measures of turbulence may advantageously be expressed in the form $\xi$ in

$$\frac{\partial \chi}{\partial t} = \frac{\partial}{\partial p}\left(\xi \frac{\partial \chi}{\partial p}\right),$$

where $p$ is the pressure (here used as a measure of height), $t$ the time, and $\chi$ may be horizontal velocity in a fixed azimuth, or potential temperature, or water per mass of atmosphere. It is suggested that $\xi$ might be called the "turbulivity." Its dimensions are grm.$^2$ cm.$^{-2}$ sec.$^{-5}$. Better still is the conductivity $c = \xi \rho^{-1} g^{-2}$.

In Part II. the eddy-shearing stress on the ground is deduced from pilot balloon observations. Values on land in any self-consistent dynamical units are found to range from 0·0007 to 0·007 times the value of $\overline{m}^2/\rho$, where $\overline{m}$ is the mean momentum per volume up to a height of 2 km. and $\rho$ is the density. Compare G. I. TAYLOR, 'Roy. Soc. Proc.,' A, vol. 92.

In Part III. evidence is given to show that the eddy-viscosity across the wind at Lindenberg increases with height, and, except near the ground, is much greater than the eddy-viscosity along the wind. Here $\xi$ ranges from $10^4$ to $5 \times 10^5$.

In Part IV. the spreading of a lamina of smoke is considered. Values of $\xi$ ranging from 7 to 140,000 are found. $\xi$ increases both with height and with velocity.

In Part V. the derivation of $\xi$ from smoke observations is examined more thoroughly.

Part VI. deals with OSBORNE REYNOLDS' eddy-stresses. For one occasion an attempt was made to measure simultaneously all six components of stress by observing the motion of thistledown. The three direct stresses are easily measured. Not so the shearing stresses however, one was found to be 2·4 times its probable error.

Part VII. summarizes the theory of scattering of particles.

Part VIII. contains numerical values derived from scattering.

In Part IX. the turbulivity $\xi$ is estimated from the rising of cumuli in calm weather and found to be $10^6$, applicable only in the sense of friction. Thus the whole range of $\xi$ observed in the free atmosphere was from 7 to a million in contrast with 0·2 in perfectly still air in a laboratory. The eddy-stresses observed have ranged in absolute value from 0·004 to 110 dynes cm.$^{-2}$.

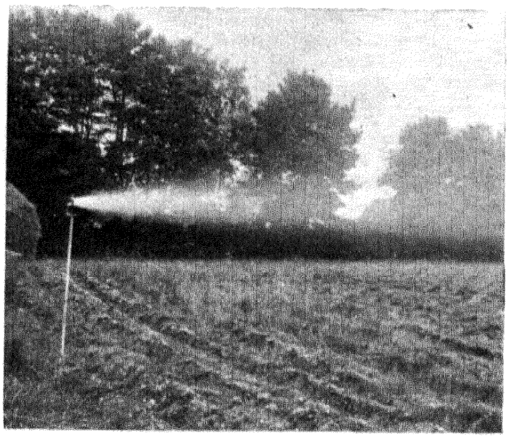
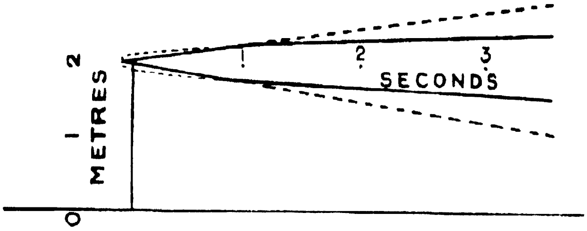


Fig. 1.



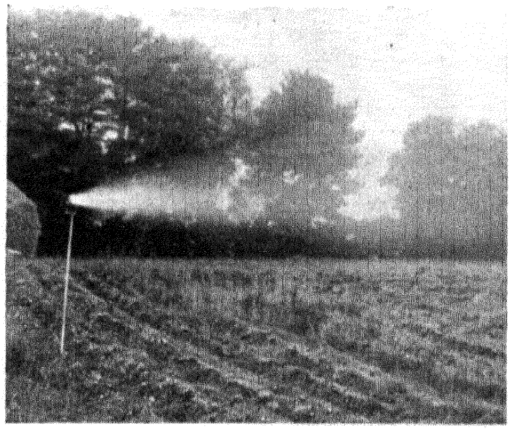
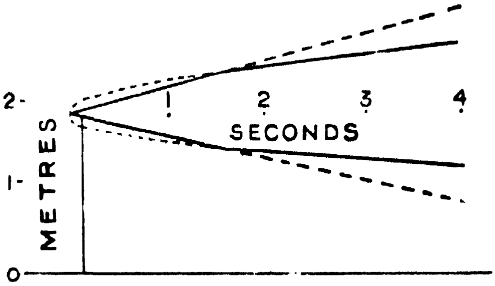


Fig. 2.



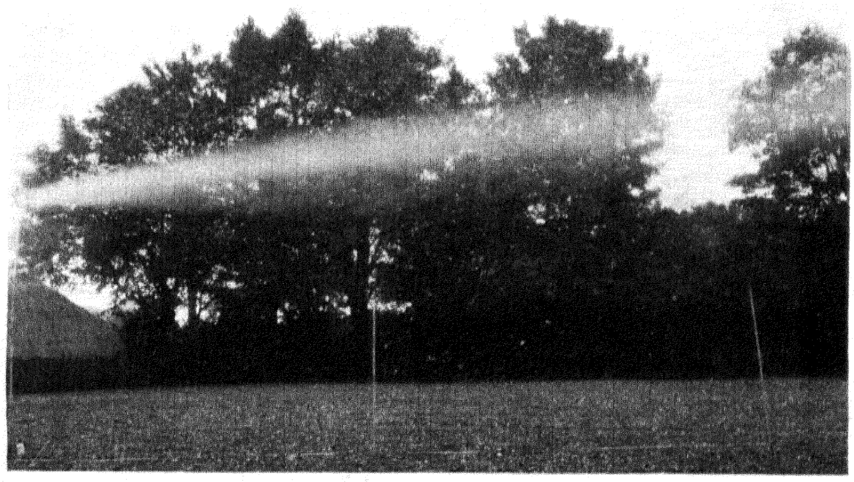


Fig. 3.

# II. *On a Theory of the Second Order Longitudinal Spherical Aberration for a Symmetrical Optical System.*

*By* T. Y. BAKER, *B.A., Instructor Commander, R.N., and* L. N. G. FILON, *M.A., D.Sc., F.R.S., Goldsmid Professor of Applied Mathematics and Mechanics in the University of London.*

Received December 2, 1919,—Read February 12, 1920.

## § 1. *Statement of the Problem and Historical References.*

IF we consider a pencil of rays issuing from a point on the axis of a symmetrical optical system (*i.e.*, a system of refracting spherical surfaces, the centres of which lie on a straight line called the axis of the system), it is well known that, if the pencil be a thin one, of which the mean ray is along the axis, the first approximation to the emergent pencil is another punctual pencil, of which the rays pass through an image point, also situated on the axis. The general method of treatment of such image points, which are usually referred to as "geometrical" images, is due to GAUSS, and is developed in any text book of Geometrical Optics.

When, however, the pencil considered is one of finite aperture, the outlying rays do not, after emergence, pass through the Gaussian image point, nor do they have the inclination assigned to them by the Gaussian calculation. The emergent rays lying in any one axial plane touch an envelope or caustic, which has one cusp at the Gaussian image, with the axis as proper tangent. The intercepts of any given emergent ray upon the axis and the image plane, measured from the Gaussian image, are known as the longitudinal and transverse spherical aberrations of that ray.

It is clear that if both these spherical aberrations, or either of them together with the inclination of the ray on emergence, be known for every possible position of object and image, and for every possible inclination of the incident ray, the whole complex of emergent rays lying in axial planes can be mapped out. The calculation of these aberrations is therefore of fundamental importance in practical optical design, where we do not deal with infinitely thin pencils.

The method employed hitherto for dealing with aberrations from the mathematical standpoint has been to develop the sines occurring in the refraction equations at each spherical surface in ascending power of some argument, which may be either the circular measure, or the sine, or the tangent, of one of the angles concerned, and to calculate, by the usual methods of successive approximation, the required aberrations as a series of ascending powers of such argument.

[Published June 16, 1920.

When this is done it is found that the terms due to the first power of the argument lead to the Gaussian image point, so that the series begin with a term involving the second power of the argument in the case of the longitudinal aberration, and the third power in the case of the transverse aberration. These terms are the first order. The following terms next in sequence, which are of fourth and fifth power respectively, are spoken of as aberrations of the second order, and so on.

A considerable amount of theoretical work has been done on aberrations of the first order by SEIDEL, ABBE and others, and the treatment of these is fairly well known. Unfortunately, it is found in practice that the first order aberrations do not give a sufficient approximation for the optician's requirements. In fact for a certain range of object and image positions, they are so badly out that they cannot be said to constitute an approximation at all. This fact has long been recognised by optical designers, whose practice is invariably to calculate, using the exact trigonometrical equations which involve no approximation at all, the correct paths of a number of selected rays, from which they draw conclusions as to the efficiency, or otherwise, of the proposed system from the practical point of view.

The trigonometrical method, however, from the designer's point of view, has the radical defect that, while it gives partial information about the performance of a given system, it gives no direct intimation of the direction in which the elimination of various defects is to be looked for, and it entails a long and laborious process of seeking for the optimum by trial and error.

The object of the authors of the present paper has been to develop a method of expressing the aberrations, which, while carrying the algebraic development to a stage including the second order, should be free from certain grave troubles involved by failure of convergency, troubles which appear to have been hitherto neglected. In fact this method gives numerical results that, for a single lens, are considerably more accurate than the ordinary second order formulæ. Further, these methods enable one to deal, in a comparatively easier form, with the problem of the second order aberrations of combinations of surfaces and systems, a problem which, so far as we know, has never been attacked from any general standpoint. KOENIG and VON ROHR (VON ROHR, 'Theorie der Optischen Instrumente,' Cap. V.) give a development of a formula for the coefficients of first order and second order in the longitudinal spherical aberration, based on ABBE's method of Invariants, but so far as can be seen, no definite results are obtained for the second order terms.

DENNIS TAYLOR ('System of Applied Optics,' p. 67) gives a formula for the spherical aberration, developed in powers of the intercept made by the ray on the first principal plane, which includes terms of second order. But his formula, a particular case of those dealt with in the present paper, is limited to the thin lens, and no attempt seems to be made at anything like a general treatment of such aberrations.

Another important object of the method to be described is to express the aberrations in such a form that, in a combination of surfaces and lenses, the effect of a given surface or lens on the final result can be readily traced. This is fundamental for the designer, who usually proceeds to sketch out his system by Gaussian methods only, being guided therein by considerations of magnification, illumination, and field of view; and then goes on to eliminate the resulting image defects, so far as he can, by bending the lenses, *i.e.*, by altering their mean curvature without changing the focal length. In doing this he usually corrects one defect at a time, with the frequent result that, when, having corrected one defect by means of one lens, he proceeds to correct a second defect, he thereby causes the reappearance of the first.

If the effects of any given lens, however, are made apparent in the final formula, it becomes a more manageable problem to devise variations which will keep any one defect *invariant* whilst others are being dealt with.

## § 2. *Notation.*

There is no general agreement among mathematical writers as to the notation employed in dealing with optical problems, and it will be convenient to state here the symbols we have adopted. They are a modification of a system due to STEINHEIL.

The successive media, proceeding in the direction of travel of the light (from left to right in our figures), are denoted by even suffixes 0, 2, 4, &c., and the same suffixes affect the rays in these media, their inclinations, $\alpha_0$, $\alpha_2$, $\alpha_4$, &c., to the axis, and their intersections $I_0$, $I_2$, $I_4$, &c., with that axis.

The successive geometrical images will be denoted by the letter J, thus $J_0$, $J_2$, $J_4$, &c. The successive surfaces of separation will be denoted by the odd suffixes 1, 3, 5, &c., and the same suffixes will affect the centres of curvature, the intersections of rays with the surfaces, and the points where the axis crosses the surfaces. The latter will be denoted by the letter A and the centres of curvature by the letter C.

Fig. 1 illustrates the use of this notation for two refracting surfaces.

The radii of curvature are $r_1$, $r_3$, $r_5$, &c., and are to be considered positive when $A_{2n+1}\,C_{2n+1}$ is measured from left to right.

The perpendicular from a centre of curvature on a ray is denoted by $p$ and is affected by a double suffix, the first belonging to the centre of the curvature and the second to the ray. Thus $p_{12}$ is the perpendicular from the centre of curvature $C_1$ of the first reflecting surface upon the ray in the second medium. Where there is no ambiguity the first suffix will usually be omitted.

The refractive index will be denoted by $n$ and affected by the suffix of its medium.

Transverse magnifications will be denoted by M. The magnification produced by surface 1 will be denoted, as convenient, by $M_1$ or $M_{02}$; by surfaces 1, 3 combined either by $M_{13}$ or $M_{04}$: by surfaces 1, 3, 5, combined either by $M_{135}$, or $M_{06}$, and so on.

The advantage of the double even suffix notation in this case is that we have a symbol, $M_{40}$, for the magnification when light passes backwards through the system, the order of the suffixes being material. Where odd suffixes are used, we have to use $1/M_1$, $1/M_{13}$, &c., for the reversed magnifications.

Ray magnifications will be denoted by $\mathbf{M}$. These are the limit of the sine-ratio for small inclinations, thus $\mathbf{M}_1 = \mathbf{M}_{02} = \underset{\alpha_0 \to 0}{\mathrm{L}} \sin \alpha_0 / \sin \alpha_2$. $M = \mathbf{M}$ when the initial and final media are the same.

With regard to inclinations, they will be treated as positive when the rays *converge* to the axis, as in fig. 1. The inclinations of the rays calculated by Gauss' process will be denoted by $\beta$. Thus $\beta_0 = \alpha_0$, $\tan \beta_2 = \tan \alpha_0/\mathbf{M}_{02}$, $\tan \beta_4 = \tan \alpha_4/\mathbf{M}_{04}$, &c.

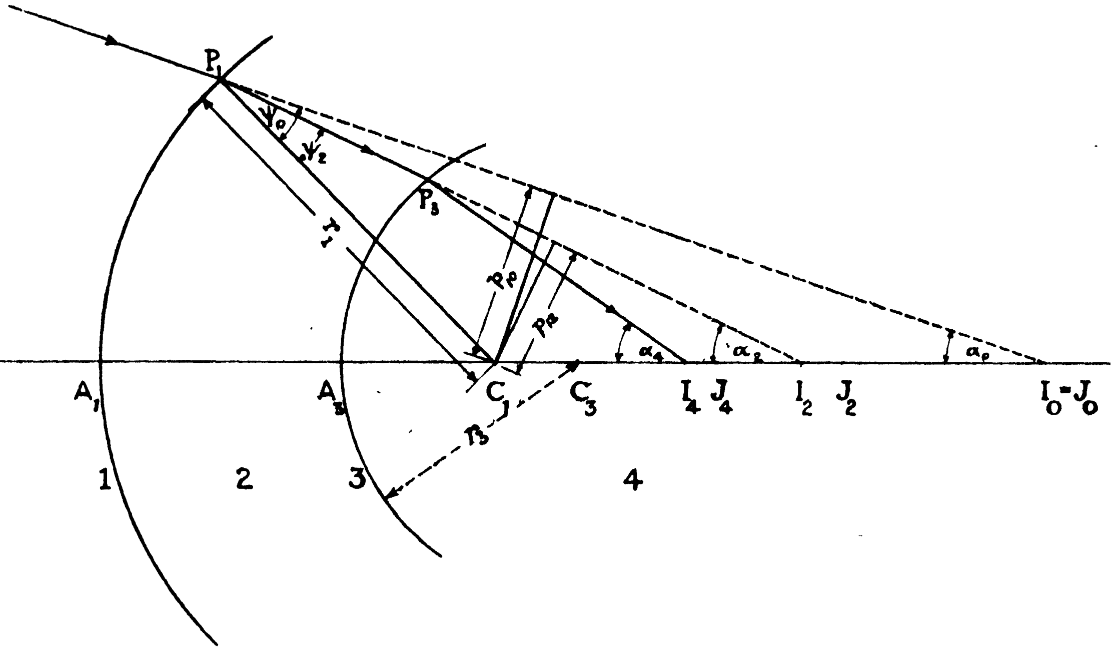


Fig. 1.

We may also use angles $\gamma$, calculated from a constant sine ratio, viz., $\gamma_0 = \alpha_0$, $\sin \gamma_2 = \sin \alpha_0/\mathbf{M}_{02}$, $\sin \gamma_4 = \sin \alpha_0/\mathbf{M}_{04}$, &c.

Throughout much of the work we shall use the same trigonometrical function (tangent or sine) of the angles $\alpha$. If the tangent is used, we shall employ the following abbreviations :—

$$q_{2n} = \tan \alpha_{2n} \qquad t_{2n} = \tan \beta_{2n}.$$

If the sine is used, the meaning of $q$, $t$ will be as follows :—

$$q_{2n} = \sin \alpha_{2n} \qquad t_{2n} = \sin \gamma_{2n}.$$

It will be found that many formulæ remain unaltered, whichever of the two interpretations for $q$ and $t$ is used.

All distances parallel to the axis will be denoted by $x$ (the attribution of the symbol being indicated in each case) and will be measured positively from left to right.

The longitudinal aberration will be denoted by $\Delta x$, with the suffix of the medium, and will be reckoned positive when the point of intersection of the ray with the axis is to the right of the geometrical image. Thus $\Delta x_2 = J_2I_2$. This is opposite to the usual convention which is based on the fact that for positive or convergent lenses, $I_2$ is generally to the left of $J_2$, ; but, in the first place, this is not universally true, and, in the second place, the convention adopted by us was found more convenient in handling the algebra.

It is to be noted that, with the notation used, the well-known formula for a lens $\frac{1}{u}+\frac{1}{v}=\frac{1}{f}$ becomes $\frac{1}{v}-\frac{1}{u}=\frac{1}{f}$, the distances $u$ and $v$ being measured in the same direction.

The distances between successive refracting surfaces we denote by $c$, with the suffix of the medium.

In dealing with a system, especially where the initial and final media are not the same, it is very convenient to use an "equivalent" Gaussian system, in which lengths parallel to the axis are measured in each medium in terms of a unit proportional to its absolute refractive index.

If we denote the corresponding points in the equivalent Gaussian system by accents, we find that

$$A'_1J'_0 = \frac{A_1J_0}{n_0}, \qquad A'_1J'_2 = \frac{A_1J_2}{n_2}, \qquad A'_1A'_3 = c'_2 = \frac{c_2}{n_2},$$

$$A'_3J'_2 = \frac{A_3J_2}{n_2}, \qquad A'_3J'_4 = \frac{A_3J_4}{n_4}, \text{ \&c.}$$

If then we denote the quantities $\frac{n_2-n_0}{r_1}$, $\frac{n_4-n_2}{r_3}$, &c., by $\frac{1}{f_1}$, $\frac{1}{f_3}$ ...; $f_1, f_3$ ... may be called the focal lengths of the successive refracting surfaces. The equations connecting image and object in the equivalent Gaussian system are

$$\frac{1}{A'_1J'_2}-\frac{1}{A'_1J'_0}=\frac{1}{f_1},$$

which is of the same form as the equation connecting image and object for a thin lens at $A'_1$.

Now bearing in mind that $A_3J_2 = A_1J_2 - A_1A_3$ and therefore $A'_3J'_2 = A'_1J'_2 - A'_1A'_3$, it is easy to show that the effects of the successive refracting surfaces in the actual system can be obtained by compounding a corresponding set of thin lenses in the equivalent Gaussian system. By dealing with the latter, we get rid of the asymmetry introduced by the difference of initial and final index. Of course this applies only to the calculation of the geometrical images. We note that in the equivalent Gaussian system the ray and transverse magnifications are identical and agree with the transverse magnifications in the actual system.

Finally the intercept of the incident ray on the leading principal plane of a system will usually be denoted by $y$. This is taken by some authors as the argument of the development in series, but differs only by a factor from $\tan \alpha_0$ or $\tan \beta_{2n}$.

In many cases it will be convenient, in order to avoid unnecessarily large suffixes, to *condense* a system of surfaces or lenses, affecting quantities referring to the system itself with suffix 1, and the initial and final media with suffixes 0 and 2, the paths in the intermediate media not being explicitly considered.

### § 3. *Singularities and Convergency.*

Consider any symmetrical optical system, of which PL and QM (fig. 2) are the initial and final refracting surfaces. Let $F_0$ be the front focus of the system and $UF_0V$ the caustic for backward-travelling rays which are parallel in the final medium.

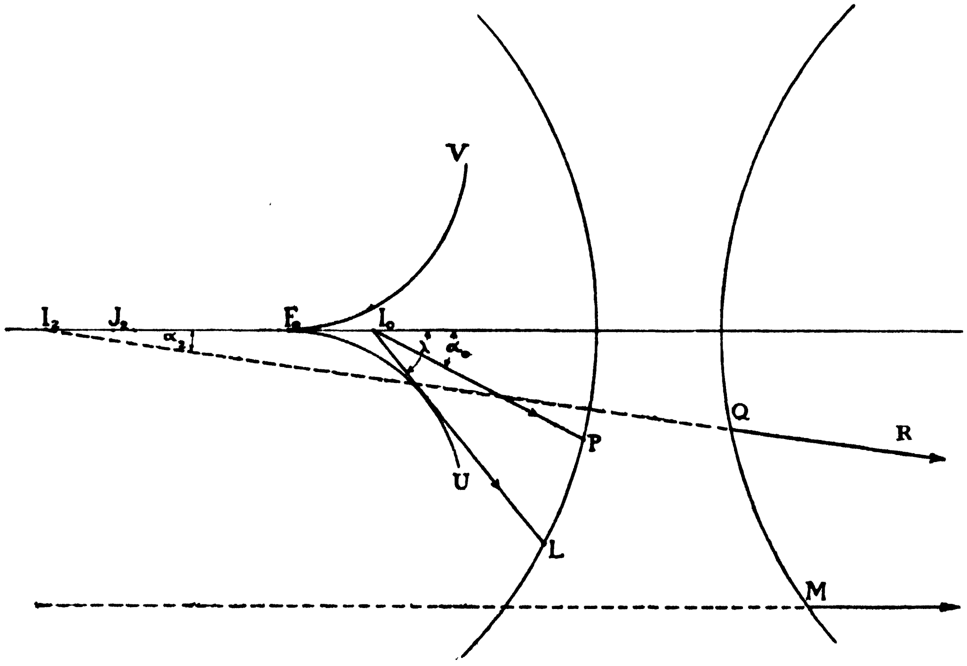


Fig. 2.

This caustic, as is well known, will usually be of the type shown in fig. 2, approximating to a semi-cubical parabola with a cusp at $F_0$, and, to fix ideas, we shall suppose the point of the cusp to be turned to the left. In the opposite case, an obvious modification of the argument will be found to lead to similar conclusions.

Any ray in the initial medium, which touches this caustic, must emerge parallel to the axis after passing through the system.

Let $I_0$ be an object point on the axis behind $F_0$ and sufficiently near to it for a real tangent to be drawn from $I_0$ to the caustic and yet go through the system. This will involve for $I_0$ a positive ray magnification $M$ exceeding some definite *finite* limit. The Gaussian image point $J_2$ will then lie a finite distance in front of $F_0$.

Consider first a nearly paraxial incident ray $I_0P$. Such a ray will be refracted approximately according to the Gaussian law and will emerge at an inclination $\alpha_2$, where $\alpha_2$ is nearly equal to $\alpha_0/M$ and both $\alpha_0$ and $M$ being finite and positive, $\alpha_2$ is also finite and positive. The ray emerges as QR, passing through a point $I_2$, finitely different from $J_2$.

As $\alpha_0$ increases, $\alpha_2$ at first increases with it, but as $\alpha_0$ reaches the value $\lambda$, corresponding to the inclination of the ray $I_0L$ which touches the front focus caustic, $\alpha_2$ is again zero. Hence between those two values $\alpha_2$ has at least one *maximum*, and for a given value of $\alpha_2$ there are at least two values of $\alpha_0$.

Thus, *within the range of values which are of practical importance*, $\alpha_0$ is a many-valued function of $\alpha_2$ having one or more branch-points, of which the one of least modulus corresponds to the first maximum of $\alpha_2$.

Now, within the same range of values, all the aberrations must be given as single valued functions of $\alpha_0$, since clearly there can only be one physical emergent ray, corresponding to one given physical incident ray. This statement, as we shall see, needs to be qualified when we are dealing with purely geometrical rays, but this need not affect the present stage of the discussion.

In consequence, if any aberration be expressed in terms of $\alpha_2$—or of any trigonometrical function of $\alpha_2$—that aberration must, in general, be a many-valued function of $\alpha_2$, having for its branch-point of least modulus the first maximum value of $\alpha_2$ mentioned above. It follows by a well-known result in theory of functions, that no TAYLOR'S series in $\alpha_2$, or in $\sin \alpha_2$, or $\tan \alpha_2$, can be valid for values of $\alpha_2$ exceeding this modulus numerically. For such values the series will be definitely divergent.

It is interesting to consider what happens when $I_0$ is on the other side of $F_0$, so that we are dealing with a large negative magnification. In this case no real tangent can be drawn from $I_0$ to the front focus caustic and the value of $\alpha_0$, for which $\alpha_2 = 0$, is a pure imaginary. But here again, although we are now dealing with imaginary values, we get two values of $\alpha_0$ for a given (pure imaginary) value of $\alpha_2$, and, although no such maximum of $\alpha_0$ occurs in the purely real values, the modulus of the imaginary branch-point limits the validity of TAYLOR'S series in $\alpha_2$ as before.

Thus there exists always a certain range, extending a finite distance (depending on the nature of the optical system) on either side of the front focus, within which no development of any aberration in powers of $\alpha_2$ or of its trigonometrical functions (or, indeed, by similar reasoning, of any inclination of the ray, except in the original medium) is valid for the whole pencil of rays which actually traverse the system.

Indeed, as the object point $I_0$ approaches the front focus, it is clear that both $\lambda$ and the maximum $\alpha_2$ tend to zero, so that only an infinitesimal portion of the rays can be dealt with by the method of successive aberrations, *i.e.*, by the TAYLOR'S series.

That the range of failure is by no means an unimportant one is shown by an example given by the authors in a paper read before the Optical Society in December, 1918. In this example the system considered is a positive lens of unit focal length and thickness $\frac{1}{16}$, meniscus shaped, with curvatures 1 and 2·36, and its convex side towards the incoming light. For such a lens and magnification as low as 2, the critical value of $\alpha_4$ is found to be about 4° 40′, corresponding to a value of $\alpha_0$ of 13°, whilst the greatest practical value of $\alpha_0$ is 26°, so that in this case only about $\frac{1}{4}$ of the light going through the lens could be dealt with by series in terms of the emergent angle. From $M = 2$ to $M = \infty$ the conditions are still worse.

As a matter of fact, it appears that in this case the range of magnifications, within which development in terms of the emergent inclinations is possible for all rays travelling through the lens, is restricted to a range lying somewhere between $M = -1$ and $M = 1·5$. This makes it clear that we cannot depend, in the calculation of the aberrations of an optical system, upon any series with the emergent inclination as argument. This is important, because from other considerations it would have been valuable to have been able to express the equation of the emergent ray in the form

$$y + qx = f(q)$$

where $q$ is the inclination of the emergent ray, and to proceed to obtain successive approximations to the caustic by developing $f(q)$ in powers. It now appears that this is not, in general, legitimate.

We now come to the consideration of series proceeding by powers of $\alpha_0$, or of its trigonometrical functions. Here the question of many-valuedness will not occur, except as follows.

If we consider a ray impinging upon a spherical refracting surface, this ray, if produced, will meet the surface at a second point. Treating the problem from the purely analytical standpoint, this second point is also one at which refraction takes place, and thus, for the same $\alpha_0$, there will, in general, be two values of $\alpha_2$, four of $\alpha_4$, and so on. $\alpha_{2n}$ will therefore, in general, be a multiple-valued function of $\alpha_0$, and the aberrations will also be multiple-valued functions, and the branch-points of these multiple-valued functions will, as before, limit the convergency of the TAYLOR series.

Now clearly two branches coincide whenever there occurs a grazing incidence; and, therefore, if the system be so arranged (as it almost necessarily is) so that no grazing incidence is reached, there will be no *real* branch-points within the range of practical values. But this does not mean that the TAYLOR'S series will necessarily be valid, for there might be imaginary branch-points. A very simple example will show how such branch-points can occur.

If $I_0$ be a source of light placed in front of a plate of thickness $c_2$ and refractive index $n$, the perpendicular from $I_0$ on the plate being the axis of the system, it is easily verified that the longitudinal spherical aberration

$$J_4I_4 = c_2\left(\frac{1}{n} - \frac{\tan\alpha_2}{\tan\alpha_0}\right),$$

where $\sin\alpha_0 = n\sin\alpha_2$, so that

$$J_4I_4 = \frac{c_2}{n}\left(1 - \frac{\sqrt{1-\sin^2\alpha_0}}{\sqrt{1-\dfrac{\sin^2\alpha_0}{n^2}}}\right).$$

The branch-points here correspond to $\alpha_0 = \frac{1}{2}\pi$ or $\alpha_2 = \frac{1}{2}\pi$, *i.e.*, to grazing incidence at the first or second surface respectively.

Clearly if $n > 1$, then, since $\sin^2\alpha_0 \leq 1$, the second grazing incidence can never occur for real values of $\alpha_0$.

But if we take as our argument $t_0 = \tan\alpha_0$, which removes the first branch-point to infinity, we find

$$J_4I_4 = \frac{c_2}{n}\left(1 - \frac{1}{\sqrt{1+t_0^2\left(1-\dfrac{1}{n^2}\right)}}\right),$$

and this has imaginary branch-points where $t_0 = \pm\dfrac{in}{\sqrt{(n^2-1)}}$. The radius of convergence of the TAYLOR's series in $t_0$ is therefore given by $t_0 = \dfrac{n}{\sqrt{(n^2-1)}}$, a value which does not correspond to any physical limitation of the rays. This applies to both the longitudinal and the transverse spherical aberrations in this case.

The above example also brings out another important point; for if in it $\sin\alpha_0$ is taken as the argument, the branch-points are $\pm 1$, $\pm n$; both of which correspond to definite physical limitations, viz., grazing incidence and total internal reflection, so that in this case the limitations of the TAYLOR's series are also the limitations of the problem.

We see then that the validity even of the expansion in $\alpha_0$ may be limited by the existence of branch-points, and that the choice of the particular trigonometrical function in which we expand may exercise a considerable influence on the result.

The limitation of the $\alpha_0$ developments due to branch-points will not, however, as in the case of the $\alpha_2$ developments, lead to vanishing radii of convergence. There is always a finite region within which these developments may be used. In what follows, therefore, we have exclusively used $\alpha_0$ as argument.

In dealing with the longitudinal spherical aberration another limitation presents itself. We have seen that if $\alpha_0 = \lambda$ (fig. 2), $\alpha_2 = 0$. It follows that the intersection of the emergent ray with the axis is then at infinity, or the longitudinal aberration is infinite and afterwards changes sign. Thus the values $\alpha_0 = \pm\lambda$ correspond to poles of the longitudinal aberration. These poles, being the singularities of least modulus, govern the convergence of the TAYLOR's series in this neighbourhood, and this radius of convergence tends to zero as the object approaches the front focus. It was a consideration of this difficulty which primarily led us to put the longitudinal aberration in a new form.

This difficulty does not arise with the transverse spherical aberration. The poles of the longitudinal spherical aberration are due to the zeros of $\alpha_0$, and on multiplying by $\frac{\tan\alpha_0}{\mathbf{M}}$ to get the transverse aberration, these poles disappear.

## § 4. *Summary of Method and Results.*

The general principle of the method employed was suggested by an attempt to fit an empirical formula to the longitudinal aberration of a lens for a certain range of curvatures and object and image positions. This empirical formula was discussed by the authors in a paper recently read before the Optical Society* and was found to give, on the whole, a singularly good fit. Briefly stated, the formula is of the following type:—

$$\Delta x = \frac{\mathrm{A}t^2}{1+\mathrm{B}t^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

where $t$ is the slope of the emergent "Gaussian" ray, so that $t = t_0/\mathbf{M}$. A is the (known) theoretical constant of the first-order aberration, which is a quartic in the magnification, and B is a cubic in the magnification, the coefficients in which are determined empirically. This formula was found to give a good approximation, even when the magnification was high and we were working well outside the limits of convergency of the TAYLOR's series for $\Delta x$.

If we consider any given object point, the longitudinal spherical aberration will be a function of $t_0$, that is, of $t$. Denoting it by $f(t)$, the reasoning of the preceding section shows that $f(t)$ is always one-valued for a finite (and generally quite considerable) range of $t$, but it is not regular, having poles at $t = \pm\tau$, where $\tau = \tan\lambda/\mathbf{M}$ and becomes rapidly small as the magnification increases numerically.

If, however, we write $(1-t^2/\tau^2)f(t) \equiv \phi(t)$, $\phi(t)$ is now limited only by the original branch-points of $f(t)$ and will, in general, have an adequate radius of convergence. We may therefore expand it in a TAYLOR's series, and we get for $f(t)$ the form

$$\Delta x = f(t) = \frac{at^2+bt^4+ct+\ldots}{1-t^2/\tau^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (2)$$

---

* BAKER and FILON, "On an Empirical Formula for the Longitudinal Spherical Aberrations in a Thick Lens," 'Proceedings of the Optical Society,' December, 1918.

If the series in the numerator converges rapidly, it will be sufficient, provided $a$ is not near zero, to stop at the first term, and we get as an approximate formula

$$\Delta x = \frac{at^2}{1-t^2/\tau^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (3)$$

which is of the same form as (1).

We have necessarily $a = \mathrm{A}$, and if the two formulæ are to tally we should have in addition $\mathrm{B} = -1/\tau^2$.

The formula in the form (3), however, is not rigorously correct to the second order of aberrations inclusive, unless $b$ happens to be small. If we wish to retain second order terms complete, we have to use

$$\Delta x = \frac{at^2+bt^4}{1-t^2/\tau^2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

and this can be written, to the same order of *algebraic* approximation, in the form

$$\Delta x = at^2/\{1-(1/\tau^2+b/a)\,t^2\}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

provided again $a$ is not zero.

If this form (5) is adopted, then the B of the empirical formula should be $1/\tau^2+b/a$. But if this is done, the formula suffers from two defects: (i) it fails whenever $a$ is near zero; (ii) it does not give exact compensation for the poles in the critical range for M large.

The further question then arose: how far are formulæ of type (4) or (5) suitable for dealing with *combinations* of surfaces or lenses? An important guiding consideration, in all work of this kind, must be the relative simplicity of the formulæ in passing from a single surface or lens to a combination, and whether these formulæ are suitable for tracing the effect of individual surfaces or lenses upon the final result.

We have ultimately been led to the conclusion that no *single* formula can satisfy completely the three ideal requirements, viz.: (i) exact agreement with development as far as the second order inclusive; (ii) simplicity in dealing with combinations; (iii) exact compensation of the poles in the critical range of M.

The method finally adopted satisfies conditions (i) and (ii). It only satisfies (iii) approximately. Numerical calculations show that numerically the approximation is adequate in the case of a lens or a simple surface. In the case of more complicated systems we have, as yet, no numerical data.

The first part of the investigation deals with the single refraction. It is there shown that the longitudinal aberration can be put into the form (1), *i.e.*, $\Delta x = \mathrm{A}t^2/(1+\mathrm{B}t^2)$, where the formula is correct to the second order inclusive.

We also find, for the inclination of the emergent ray, the formula

$$q = t\,(1+\mathrm{B}t^2)/(1+\mathrm{C}t^2), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (6)$$

which is correct to the first order when $q$ and $t$ are tangents and to the second order when $q$ and $t$ are sines.

In the above A, B, C are polynomials in M of degrees 4, 3, 2 respectively, so that the empirical formula is well justified for the simple refracting surface. In this case, too, it is possible to calculate $\tau$ directly, and, in fact, a simple geometrical construction is given for it. When this is followed for varying image-positions, it is found that outside a certain range of M, the $\tau$ so obtained becomes irrelevant, and that, in fact, if the correct factor $1-t^2/\tau^2$ is retained in the denominator of $\Delta x$, although it improves the fit by removing singularities in the range round $M = \infty$, it introduces entirely fictitious singularities in other and important parts of the range, and makes the formula worthless.

A good deal of light is thrown upon the problem when it is found that, if we develop $\frac{1}{\tau^2}$ in descending powers of M in the neighbourhood of $M = \infty$, the two leading terms are discovered to be identical with the two leading terms of the cubic B, previously obtained. This makes our B approximate more and more closely to $\frac{1}{\tau^2}$ precisely as the effect of the denominator term becomes more important, and it is this fact which is the key to the numerical value of the method.

We then proceed to show how the constants for a combination of the two systems can be obtained from the corresponding constants of the individual systems. In doing this it appears that, so soon as we pass from the single refracting surface to the lens, a new constant is introduced into the formula, which now takes the form

$$\Delta x = \frac{At^2 + Et^4}{1 + Bt^2} \quad \dots\dots\dots\dots \quad (7)$$

where A and B are of the same form as before, but E is now a polynomial of degree 6 in M. In the case of a lens the term $Et^4$ is found to be, in general, of small importance, which accounts for the good fit of the empirical formula.

The formula (7) *for a combination* holds good to the second order inclusive, and B agrees with $\frac{1}{\tau^2}$, when M is large, as far as the leading term only. For numerical purposes, however, a correction is discussed, which is very readily applied, and which makes the two leading terms in B agree with the two leading terms in $\frac{1}{\tau^2}$, as in the case of the single refracting surfaces.

The formulæ for combining two systems take comparatively simple forms; the A, B and C for the combination are expressed as linear functions of the A's, B's and C's of the components, and the E as a lineo-linear function of the A's, B's and C's of the components, each term involving a product of which one factor belongs to one component of the combination and the other factor to the second. In addition the E for the combination involves linear terms in the E's of the components.

These results are found to hold good in the more general case of the combination of three or more systems. It will follow that if the constants A, B, C, E are tabulated for lenses of various curvatures, the effect of the combination can be traced relatively easily and the aberrations corrected, so far as possible, by suitably bending the lenses, while keeping the general arrangement and the magnifications the same.

Explicit values of the constants for the single refracting surface and a single thick or thin lens have been obtained and are tabulated for reference, so as to be available for eventual computation of the required tables. We have also given some numerical values for a single lens, and a numerical test of the accuracy in this case, which works out at about $\frac{1}{500}$ of the total aberration for the range of cases taken.

The corresponding formulæ with sin $\gamma$ instead of tan $\beta$ as argument are discussed, and it is shown that the equations of combination are of the same form as before.

Certain invariant relations between the coefficients in A, B, C, E are developed, which enable various calculations to be simplified and in particular to determine these constants for a system reversed, when they are known for the direct system. This will generally halve the work of tabulation.

## § 5. *The Single Refracting Surface.*

Using the general notation described in § 2, consider refractions at a single refracting surface.

Let $\psi_0$, $\psi_2$ denote the angles of incidence and refraction, so that $\psi_0 = \mathrm{C_1P_1I_0}$, $\psi_2 = \mathrm{C_1P_1I_2}$ (fig. 1).

Let $\mathrm{C_1I_0} = \mathrm{X_0} = x_0$, $\mathrm{C_1I_2} = \mathrm{X_2} = x_2 + \Delta x_2$, $\mathrm{C_1J_2} = x_2$ we then have the set of refraction equations

$$\sin\psi_0 = p_0/r_1 = x_0 \sin\alpha_0/r_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (8)$$

$$\sin\psi_2 = p_2/r_1 = \mathrm{X_2} \sin\alpha_2/r_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (9)$$

$$n_2 p_2 = n_0 p_0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

$$\alpha_2 - \alpha_0 = \psi_0 - \psi_2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

Let $\xi_0 = \mathrm{A'_1I'_0}$, $\xi_2 = \mathrm{A'_1J'_2}$ in the "equivalent" Gaussian system (see § 2). Then

$$n_0\xi_0 = x_0 + r_1, \qquad n_2\xi_2 = x_2 + r_1, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

and we find, using the first approximation when $\alpha_0$, &c., are small

$$1/\xi_2 - 1/\xi_0 = 1/f_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

If $\mathbf{M}_1 = \underset{\alpha_0 \to 0}{\mathrm{L}}\, \alpha_0/\alpha_2$, which we shall call the *ray magnification*, we have

$$\mathbf{M}_1 = n_2\xi_2/n_0\xi_0 = n_2x_2/n_0x_0. \qquad (14)$$

The transverse magnification $\mathrm{M}_1$ is given by

$$\mathrm{M}_1 = \xi_2/\xi_0 = x_2/x_0 = n_0\mathbf{M}_1/n_2. \qquad (15)$$

We can also express (13) in another well-known form, namely,

$$n_0/x_2 - n_2/x_0 = 1/f_1, \qquad (16)$$

whence, using (14), we obtain

$$\left.\begin{aligned} x_0 &= n_2 f_1 (1-\mathbf{M}_1)/\mathbf{M}_1 \\ x_2 &= n_0 f_1 (1-\mathbf{M}_1) \end{aligned}\right\} \qquad (17)$$

Again, from (9), (8) and (10)

$$\begin{aligned} \mathrm{X}_2 = p_2/\sin\alpha_2 &= n_0x_0\sin\alpha_0/n_2\sin\alpha_2 = n_0f_1(1-\mathbf{M}_1)\sin\alpha_0/\mathbf{M}_1\sin\alpha_2 \\ &= x_2\sin\alpha_0/\mathbf{M}_1\sin\alpha_2, \qquad (18) \end{aligned}$$

and the longitudinal aberration

$$\Delta x_2 = \mathrm{X}_2 - x_2 = x_2(\sin\alpha_0/\mathbf{M}_1\sin\alpha_2 - 1). \qquad (19)$$

The corresponding longitudinal aberration in the equivalent Gaussian system is found from

$$\Delta\xi_2 = \Delta x_2/n_2 = (n_0 f_1/n_2)(1-\mathbf{M}_1)(\sin\alpha_0/\mathbf{M}_1\sin\alpha_2 - 1). \qquad (20)$$

Now from (11)

$$\begin{aligned} \sin\alpha_2 = {} & \sin\alpha_0\cos\psi_0\cos\psi_2 + \sin\psi_0\cos\alpha_0\cos\psi_2 - \sin\psi_2\cos\alpha_0\cos\psi_0 \\ & + \sin\alpha_0\sin\psi_0\sin\psi_2, \end{aligned}$$

whence, using (8), (9) and (10),

$$\begin{aligned} \sin\alpha_2/\sin\alpha_0 = {} & \{1-(x_0/r_1)^2\sin^2\alpha_0\}^{\frac{1}{2}}\{1-(n_0x_0/n_2r_1)^2\sin^2\alpha_0\}^{\frac{1}{2}} \\ & + (x_0/r_1)\{1-\sin^2\alpha_0\}^{\frac{1}{2}}\{1-(n_0x_0/n_2r_1)^2\sin^2\alpha_0\}^{\frac{1}{2}} \\ & - (n_0x_0/n_2r_1)\{1-\sin^2\alpha_0\}^{\frac{1}{2}}\{1-(x_0/r_1)^2\sin^2\alpha_0\}^{\frac{1}{2}} \\ & + (n_0x_0^2/n_2r_1^2)\sin^2\alpha_0, \end{aligned}$$

and developing this in ascending powers of $\sin\alpha_0$, we obtain, retaining only terms of fourth degree

$$\begin{aligned} \sin\alpha_2/\sin\alpha_0 = {} & 1 + x_0(n_2-n_0)/n_2r_1 - \tfrac{1}{2}\mathrm{P}\sin^2\alpha_0 \\ & - \tfrac{1}{8}\mathrm{P}\sin^4\alpha_0\{(n_2+n_0)^2x_0^2 + n_2(r_1-x_0)(n_2r_1+n_0x_0)\}/n_2^2r_1^2, \end{aligned}$$

where

$$\mathrm{P} = (1-n_0/n_2)(x_0/r_1)(1+x_0/r_1)(1-n_0x_0/n_2r_1),$$

and is the quantity whose vanishing gives the aplanatic points and must therefore be a factor of every coefficient after the first in the development of $\sin\alpha_2$ in powers of $\sin\alpha_0$.

If we write for shortness

$$Q \equiv (1+n_0/n_2)^2 x_0^2/r_1^2+(1-x_0/r_1)(1+n_0x_0/n_2r_1),$$

and remember that

$$1+x_0(n_2-n_0)/n_2r_1 = 1/\mathbf{M}_1,$$

we find

$$\begin{aligned}\sin\alpha_2/\sin\alpha_0 &= 1/\mathbf{M}_1-\tfrac{1}{2}P\sin^2\alpha_0/(1-\tfrac{1}{4}Q\sin^2\alpha_0)\\ &= \mathbf{M}_1^{-1}\{1-(\tfrac{1}{2}P\mathbf{M}_1+\tfrac{1}{4}Q)\sin^2\alpha_0\}/(1-\tfrac{1}{4}Q\sin^2\alpha_0)\\ &= \mathbf{M}_1^{-1}\{1+\mathbf{B}\sin^2\alpha_0/\mathbf{M}_1^2\}/\{1+\mathbf{C}\sin^2\alpha_0/\mathbf{M}_1^2\} \quad . \quad . \quad (21)\end{aligned}$$

correct as far as the second order inclusive, where

$$\mathbf{B} = -\tfrac{1}{2}P\mathbf{M}_1^3-\tfrac{1}{4}Q\mathbf{M}_1^2, \qquad \mathbf{C} = -\tfrac{1}{4}Q\mathbf{M}_1^2,$$

from which, after some reductions

$$\mathbf{C} = -\tfrac{1}{4}(n_2-n_0)^{-2}\{(n_2^2+n_0n_2+n_0^2)-3(n_0^2+n_2^2)\mathbf{M}_1+3(n_2^2-n_0n_2+n_0^2)\mathbf{M}_1^2\}, \quad (22)$$

and

$$\begin{aligned}\mathbf{B} &= \tfrac{1}{2}(n_2-n_0)^{-2}(1-\mathbf{M}_1)(n_2-n_0\mathbf{M}_1)(n_0-n_2\mathbf{M}_1)+\mathbf{C}\\ &= \tfrac{1}{4}(n_2-n_0)^{-2}\{-(n_2^2-n_0n_2+n_0^2)+(n_2-n_0)^2\mathbf{M}_1-(n_2^2+n_0^2-5n_0n_2)\mathbf{M}_1^2-2n_0n_2\mathbf{M}_1^3\}. (23)\end{aligned}$$

Returning to equation (19) and using (21)

$$\begin{aligned}\Delta x_2 &= x_2(\mathbf{C}-\mathbf{B})\mathbf{M}_1^{-2}\sin^2\alpha_0/\{1+\mathbf{B}\mathbf{M}_1^{-2}\sin^2\alpha_0\}\\ &= n_2f_1A\mathbf{M}_1^{-2}\sin^2\alpha_0/\{1+\mathbf{B}\mathbf{M}_1^{-2}\sin^2\alpha_0\}, \quad . \quad . \quad . \quad . \quad . \quad (24)\end{aligned}$$

where

$$\begin{aligned}A &= x_2(\mathbf{C}-\mathbf{B})/n_2f_1\\ &= -\tfrac{1}{2}(n_2-n_0)^{-2}(n_0/n_2)(1-\mathbf{M}_1)^2(n_2-n_0\mathbf{M}_1)(n_0-n_2\mathbf{M}_1)\\ &= \tfrac{1}{2}(n_2-n_0)^{-2}(n_0/n_2)\left\{\begin{array}{l}-n_0n_2+(n_2+n_0)^2\mathbf{M}_1-2(n_2^2+n_0n_2+n_0^2)\mathbf{M}_1^2\\ \quad +(n_2+n_0)^2\mathbf{M}_1^3-n_0n_2\mathbf{M}_1^4\end{array}\right\}. \quad (25)\end{aligned}$$

All the above formulæ are correct to the second order of aberrations inclusive. We note that A, **B** and **C** are polynomials of degree 4, 3 and 2 in the magnification respectively.

If we express the aberration in terms of tangents instead of sines we have at once

$$\begin{aligned}\Delta x_2 &= n_2f_1A\mathbf{M}_1^{-2}\tan^2\alpha_0/\{1+(\mathbf{B}+\mathbf{M}_1^2)\mathbf{M}_1^{-2}\tan^2\alpha_0\}\\ &= n_2f_1A\tan^2\beta_2/\{1+B\tan^2\beta_2\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (26)\end{aligned}$$

where

$$\begin{aligned} \mathrm{B} &= \mathbf{B}+\mathbf{M}_1^2 \\ &= \tfrac{1}{4}(n_2-n_0)^{-2}\{-(n_2^2-n_0n_2+n_0^2)+(n_2-n_0)^2\mathbf{M}_1+3(n_2^2-n_0n_2+n_0^2)\mathbf{M}_1^2-2n_0n_2\mathbf{M}_1^3\}. \quad (27) \end{aligned}$$

When, however, we come to develop a formula for tangents, similar to (21) for sines, it is found that, in the series for tan $\alpha_2$/tan $\alpha_0$, we do not have all the coefficients after the first vanishing together; for, even at the aplanatic points, the tangent ratio is not constant.

We can, indeed, write

$$\tan\alpha_2/\tan\alpha_0 = \mathbf{M}_1^{-1}(1+\mathrm{B}\mathbf{M}_1^{-2}\tan^2\alpha_0)/(1+\mathrm{C}\mathbf{M}_1^{-2}\tan^2\alpha_0), \quad . \quad . \quad . \quad (28)$$

and choose the coefficients B and C so that the developments shall agree as far as terms in $\tan^4\alpha_0$ inclusive. This can, in general, be done in one way only. But we then find that B and C are no longer integral functions of the magnification. Their infinities have to be taken into account, and generally the method becomes complicated and unsatisfactory.

From other considerations, however, it appears that since the zeroes of $\alpha_2$ must be the same as the poles of $\Delta x_2$, the B in (28) must be the same as the B in (26), and this will fix C as follows :—

$$\begin{aligned} \tan^2\alpha_2 &= \sin^2\alpha_2/(1-\sin^2\alpha_2) \\ &= \frac{\mathbf{M}_1^{-2}\sin^2\alpha_0\{1+\mathbf{B}\sin^2\alpha_0/\mathbf{M}_1^2\}^2}{\{1+\mathbf{C}\sin^2\alpha_0/\mathbf{M}_1^2\}^2-\mathbf{M}_1^{-2}\sin^2\alpha_0\{1+\mathbf{B}\sin^2\alpha_0/\mathbf{M}_1^2\}^2} \\ &= \frac{\mathbf{M}_1^{-2}\tan^2\alpha_0\{1+(\mathbf{B}+\mathbf{M}_1^2)\mathbf{M}_1^{-2}\tan^2\alpha_0\}^2}{(1+\tan^2\alpha_0)\{1+(\mathbf{C}+\mathbf{M}_1^2)\mathbf{M}_1^{-2}\tan^2\alpha_0\}^2-\mathbf{M}_1^{-2}\tan^2\alpha_0\{1+(\mathbf{B}+\mathbf{M}_1^2)\mathbf{M}_1^{-2}\tan^2\alpha_0\}^2}, \end{aligned}$$

substituting from (21).

Hence, taking the square root and developing the denominator in powers of tan $\beta_2$, *i.e.*, of $\mathbf{M}_1^{-1}\tan\alpha_0$,

$$\begin{aligned} \tan\alpha_2/\tan\beta_2 &= \frac{(1+\mathrm{B}\tan^2\beta_2)}{1+\tan^2\beta_2(\mathbf{C}+\frac{3}{2}\mathbf{M}_1^2-\frac{1}{2})+\frac{1}{2}\tan^4\beta_2\{\mathbf{C}(1+\mathbf{M}_1^2)+\frac{3}{4}\mathbf{M}_1^4+\frac{3}{2}\mathbf{M}_1^2-\frac{1}{4}-2\mathrm{B}\}} \\ &= (1+\mathrm{B}\tan^2\beta_2)/(1+\mathrm{C}\tan^2\beta_2+\mathrm{D}\tan^4\beta_2), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (29) \end{aligned}$$

where

$$\left.\begin{aligned} \mathrm{C} &= \mathbf{C}+\tfrac{3}{2}\mathbf{M}_1^2-\tfrac{1}{2} \\ \mathrm{D} &= \mathbf{C}(1+\mathbf{M}_1^2)+\tfrac{3}{4}\mathbf{M}_1^4+\tfrac{3}{2}\mathbf{M}_1^2-\tfrac{1}{4}-2\mathrm{B} \\ &= \mathrm{C}(1+\mathbf{M}_1^2)+\tfrac{1}{4}(1-\mathbf{M}_1^2)(1+3\mathbf{M}_1^2)-2\mathrm{B} \end{aligned}\right\}, \quad . \quad . \quad . \quad . \quad . \quad (30)$$

(29) is now correct as far as the second order inclusive.

If we only require the tangent ratio correct up to the first order of aberrations, we have the formula

$$\tan\alpha_2 = \tan\beta_2(1+\mathrm{B}\tan^2\beta_2)/(1+\mathrm{C}\tan^2\beta_2) \quad . \quad . \quad . \quad . \quad . \quad (31)$$

The value of C, when written out fully, is given by

$$C = \tfrac{3}{4}(n_2-n_0)^{-2}\{-(n_2^2-n_0n_2+n_0^2)+(n_0^2+n_2^2)\mathbf{M}_1+(n_0^2-3n_0n_2+n_2^2)\mathbf{M}_1^2\}. \quad (32)$$

### § 6. *The Convergency Factor and the Singular Inclination for a Single Refracting Surface.*

Having now obtained expressions (24 and 26) for the longitudinal spherical aberration, which are correct to the second order of aberrations, when expansion in powers of $\sin\alpha_0$ or $\tan\alpha_0$ is legitimate and rapidly convergent, we have now to enquire how far the same expression remains valid as $\mathbf{M}$ increases, in which case we know that B or **B** increases without limit and the convergency fails, even for comparatively small values of $\alpha_0$.

Here it will be convenient to introduce two definitions :—

I. We shall call *singular inclination* the value $\lambda$ (see § 3) of $\alpha_0$ for which the emergent ray is parallel to the axis.

II. The factor $1-\sin^2\alpha_0/\sin^2\lambda$ (or $1-\tan^2\alpha_0/\tan^2\lambda$ if we are dealing with tangents) we shall call the *convergency factor*. If we multiply $\Delta x$ by the convergency factor we remove those singularities of $\Delta x$ which are instrumental in causing critical failure of convergency.

To find the singular inclination and convergency factor for a single refracting surface, we have to find when $\alpha_2 = 0$.

Going back to the fundamental equations (8) to (11) we have $\alpha_2 = 0$ when $\alpha_0 = \lambda$, where

$$\lambda = \psi_2-\psi_0$$

which leads to

$$\begin{aligned}\sin\lambda &= \sin\psi_2\cos\psi_0-\sin\psi_0\cos\psi_2\\ &= (n_0x_0\sin\lambda/n_2r_1)\sqrt{\{1-(x_0\sin\lambda/r_1)^2\}}-(x_0\sin\lambda/r_1)\sqrt{\{1-(n_0x_0\sin\lambda/n_2r_1)^2\}}.\end{aligned}$$

Hence, either $\sin\lambda = 0$, which obviously refers to the axial ray, a trivial and (for our purpose) irrelevant solution, or

$$r_1/x_0 = (n_0/n_2)\sqrt{\{1-(x_0\sin\lambda/r_1)^2\}}-\sqrt{\{1-(n_0x_0\sin\lambda/n_2r_1)^2\}}. \quad (33)$$

On rationalising (33) leads to

$$\begin{aligned}4n_0^2\sin^2\lambda/n_2^2 &= 4r_1^2/x_0^2-(1+r_1^2/x_0^2-n_0^2/n_2^2)^2\\ &= -(1+r_1/x_0-n_0/n_2)(1+r_1/x_0+n_0/n_2)(1-r_1/x_0+n_0/n_2)(1-r_1/x_0-n_0/n_2). \quad (34)\end{aligned}$$

This gives the singular inclination.

If we write

$$R \equiv (1+r_1/x_0-n_0/n_2)(1+r_1/x_0+n_0/n_2)(1-r_1/x_0+n_0/n_2)(1-r_1/x_0-n_0/n_2), \quad (35)$$

it follows that

$$1+4n_0^2 \sin^2 \alpha_0/n_2^2\mathrm{R} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (36)$$

is the required convergency factor.

If the formulæ of § 5 are to get accurately over the failure of convergency, this convergency factor should be identical with

$$1+\mathbf{B} \sin^2 \alpha_0/\mathbf{M}_1^2,$$

that is, we should have

$$\mathbf{B} = 4n_0^2 \mathbf{M}_1^2/n_2^2\mathrm{R}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (37)$$

which, when written out, becomes

$$\mathbf{B} = 4n_0^2 n_2^2 \mathbf{M}_1^2 (1-\mathbf{M}_1)^4/(n_2-n_0)^2 (1-2\mathbf{M}_1)(n_2+n_0-2n_0\mathbf{M})(n_2+n_0-2n_2\mathbf{M}_1). \quad (38)$$

This does not agree with the previously found value for **B**, being of fractional form in $\mathbf{M}_1$. It does lead to **B** becoming infinite of the order $\mathbf{M}_1^3$ when $\mathbf{M}_1$ tends to infinity, but it indicates an infinity of **B** (and therefore a critical failure of convergency) at three other places, namely when $\mathbf{M}_1 = \frac{1}{2}$, $\frac{1}{2}(n_0+n_2)/n_0$, $\frac{1}{2}(n_0+n_2)/n_2$, at none of which does a failure of convergency really occur, as can readily be verified.

The reason for this is made clearer by geometrical reasoning as follows :—

Let $P_1Q_1$ (fig. 3) be a ray which is parallel to the axis in medium 2. To make the figure easier and the quantities dealt with positive, the refraction has been taken from a denser to a rarer medium, so that $n_0/n_2 > 1$.

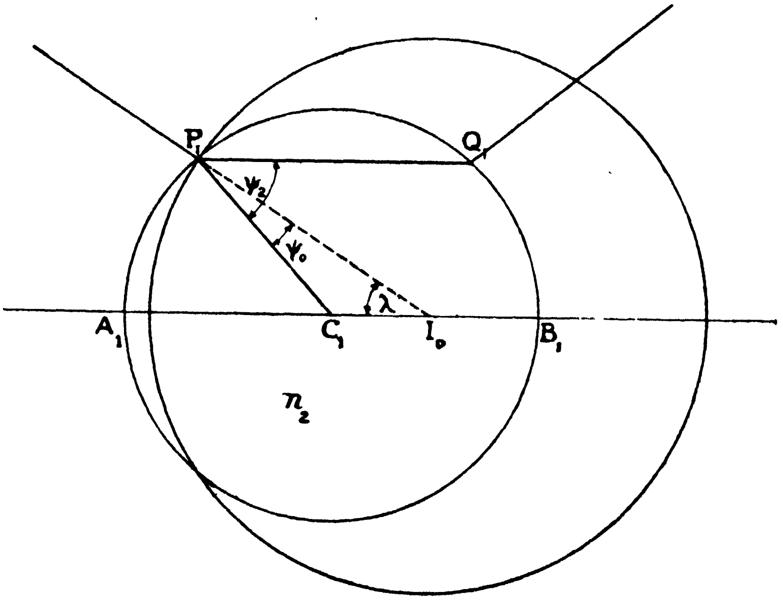


Fig. 3.

In the triangle $I_0C_1P_1$ of fig. 3 we have $\sin \psi_2/\sin \psi_0 = I_0P_1/x_0$. But

$$\sin \psi_2/\sin \psi_0 = n_0/n_2; \quad \text{hence} \quad I_0P_1 = n_0x_0/n_2.$$

The point $P_1$ and singular inclination $\lambda$ can therefore be constructed geometrically as follows.

With $I_0$ as centre and radius $n_0.C_1I_0/n_2$ describe a circle meeting the refracting surface at $P_1$; $C_1I_0P_1$ is the angle $\lambda$ required.

This angle $\lambda$ approaches zero, that is, we get a critical failure of convergency, when the two circles approach contact at $A_1$. The limiting case is, therefore, when $A_1I_0 = n_0 . C_1I_0/n_2$ or $I_0$ divides $A_1C_1$ *externally* in the ratio $n_0 : n_2$.

When this happens $r_1+x_0 = n_0x_0/n_2$ leading to $\mathbf{M}_1 = \infty$, a case of true critical failure. But clearly, by symmetry, we get a precisely similar result when $P_1Q_1$ is due to a ray entering the surface at $Q_1$, and travelling backward through medium 2. In this case the limiting position of $I_0$ divides $B_1C_1$ externally in the ratio $n_0 : n_2$ and is defined by $r_1-x_0 = -n_0x_0/n_2$ or $\mathbf{M}_1 = \frac{1}{2}$.

This case would correspond, analytically, to $\alpha_2 = \pi$, and the corresponding equation (33) would become

$$-r_1/x_0 = (n_0/n_2)\sqrt{\{1-(x_0 \sin\lambda/r_1)^2\}} - \sqrt{\{1-(n_0x_0 \sin\lambda/n_2r_1)^2\}}.$$

Now if we examine (34) we find that in the process of clearing roots, $r_1/x_0$ appears squared in the final result, which accordingly includes both $\alpha_2 = 0$ and $\alpha_2 = \pi$. If we write $r_1^2/x_0^2 = u$, then we should really write equation (33) in the form

$$\sqrt{u} = (n_0/n_2)\sqrt{(1-\sin^2\lambda/u)} - \sqrt{(1-n_0^2\sin^2\lambda/n_2^2u)}, \quad . \quad . \quad . \quad . \quad (39)$$

and the two cases are discriminated by assigning to $\sqrt{u}$ one or the other sign. One of these cases is necessarily irrelevant since refraction at the posterior surface of the sphere is physically excluded.

Further, if we consider the other two values which make $R = 0$, viz.,

$$\mathbf{M}_1 = (n_0+n_2)/2n_0 \text{ and } \mathbf{M}_1 = (n_0+n_2)/2n_2,$$

they correspond to

$$r_1+x_0 = -n_0x_0/n_2$$

and

$$r_1-x_0 = n_0x_0/n_2,$$

*i.e.*, to positions of $I_0$ in which it divides $A_1C_1$ and $B_1C_1$ *internally* in the ratio $n_0 : n_2$. But these belong geometrically to the limit of cases in which the incident and refracted rays lie on opposite sides of the normal, *i.e.*, to a negative refractive index. And indeed they are obtained from the two previous points by reversing the sign of $n_0/n_2$.

Here again, examination of (34) shows that $n_0/n_2$ appears squared in it. Therefore (34) includes the cases in question. These, however, may be obtained analytically by changing $\psi_0$, or $\psi_2$, into its supplement, *i.e.*, by reversing the sign of $\cos\psi_0$ or $\cos\psi_2$ or by changing the determination of the sign of one or other of the square roots on the right-hand side of (39).

The cases are discriminated by the vanishing of these square roots, which occurs when $\cos\psi_0$ or $\cos\psi_2 = 0$.

It appears, therefore, that the vanishing of the factors in the denominator of (38) is wholly irrelevant, and, if we adopted for **B** the value given on the right-hand side of that equation, we should thereby be introducing, in the neighbourhood of $\mathbf{M}_1 = \frac{1}{2}$, $(n_0+n_2)/2n_0$, $(n_0+n_2)/2n_2$ entirely irrelevant singularities, which would make the formula worthless.

The question arises, what is the range of values of $\mathbf{M}_1$ for which the equation

$$4n_0^2 \sin^2\lambda/n_2^2 = -\mathrm{R}$$

is valid and legitimate?

If we start from $\mathbf{M}_1 = \infty$, which corresponds to a real case, the signs of the square roots in (39) are well determined, and the correspondence between $\sin^2\lambda$ and $\mathbf{M}_1$ is unique and definite and can be continued until we reach a point where one case passes into another. These cases we have found to be the branch-points of the three square roots, namely :—

$$u = 0, \qquad \cos\psi_0 = 0 \quad \text{and} \quad \cos\psi_2 = 0.$$

$$u = 0 \quad \text{leads to} \quad x_0 = \infty \text{ or } \mathbf{M}_1 = 0 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\mathrm{A})$$

$\cos\psi_0 = 0$ leads to $\sin^2\lambda = u$, or, using the first form of (34)

$$4n_0^2 u/n_2^2 = 4u-(1+u-n_0^2/n_2^2)^2,$$

*i.e.*, $(1-u-n_0^2/n_2^2)^2 = 0$, that is $u = 1-n_0^2/n_2^2$, leading to

$$x_0 = \pm r_1\,(1-n_0^2/n_2^2)^{-\frac{1}{2}} \text{ and } \mathbf{M}_1 = \{1\pm(n_2-n_0)/\sqrt{(n_2^2-n_0^2)}\}^{-1} \quad . \quad . \quad . \quad (\mathrm{B})$$

$\cos\psi_2 = 0$ leads to $\sin^2\lambda = n_2^2 u/n_0^2$, that is, to

$$u = n_0^2/n_2^2-1,\ x_0 = \pm r_1\,(n_0^2/n_2^2-1)^{-\frac{1}{2}},\ \mathbf{M}_1 = \{1\pm(n_2-n_0)/\sqrt{(n_0^2-n_2^2)}\}^{-1} \quad . \quad . \quad (\mathrm{C})$$

If $n_0 > n_2$, both values of $\mathbf{M}_1$ given by (B) are imaginary. The values given by (C) are both positive, $\mathbf{M}_1 = \{1+(n_2-n_0)/\sqrt{(n_0^2-n_2^2)}\}^{-1}$ being the greater.

The range over which we can travel without ambiguity is, therefore, from $\mathbf{M}_1 = +\infty$ to $\mathbf{M}_1 = \{1+(n_2-n_0)/\sqrt{(n_0^2-n_2^2)}\}^{-1}$ and from $\mathbf{M}_1 = -\infty$ to $\mathbf{M}_1 = 0$.

If $n_0 < n_2$, the values of $\mathbf{M}_1$ given by (C) are imaginary, those given by (B) are positive, and $\mathbf{M}_1 = \{1-(n_2-n_0)/\sqrt{(n_2^2-n_0^2)}\}^{-1}$ is the greater, so that the range of validity is from $\mathbf{M}_1 = +\infty$ to $\mathbf{M}_1 = \{1-(n_2-n_0)/\sqrt{(n_2^2-n_0^2)}\}^{-1}$ and from $\mathbf{M}_1 = -\infty$ to $\mathbf{M}_1 = 0$.

Within this range $(1+4n_0^2 \sin^2\alpha_0/n_2^2\mathrm{R})$ is the correct convergency factor; outside this range it is irrelevant.

It is clear, then, that we cannot find a single formula for the convergency factor, which will hold for all values of the magnification.

Further, if the factor $(1+4n_0^2 \sin^2 \alpha_0/n_2^2\mathrm{R})$ is introduced into the denominator of $\Delta x_2$, we no longer obtain expressions of the simple type (24) and (26), and endless complications are introduced when we come to consider a compound system.

Can we make our expression $\mathbf{B}$ given by (23) give a tolerable approximation to $(4n_0^2\mathbf{M}_1^2/n_2^2\mathrm{R})$ for those regions where the denominator factor is really needed, namely for $\mathbf{M}_1$ large, positively or negatively?

To get the answer to this question we develop $(4n_0^2\mathbf{M}_1^2/n_2^2\mathrm{R})$ in descending powers of $\mathbf{M}_1$.

This is found to be (the most rapid method is to break up first into partial fractions)

$$\tfrac{1}{4}(n_2-n_0)^{-2}\left\{\begin{array}{l} -2n_0n_2\mathbf{M}_1^3-(n_0^2-5n_0n_2+n_2^2)\mathbf{M}_1^2 \\ -[(n_2-n_0)^4+n_0^2n_2^2]\mathbf{M}_1/2n_0n_2-[(n_2-n_0)^3(n_2^3-n_0^3)+n_2^3n_0^3]/4n_0^2n_2^2 \\ \quad +\text{terms in } 1/\mathbf{M}_1, \text{ \&c.} \end{array}\right\}. \quad (40)$$

If we now compare (40) with (23) we find that the most important terms when $\mathbf{M}_1$ is large, namely those in $\mathbf{M}_1^3$ and $\mathbf{M}_1^2$ agree in the two expressions.

We may, therefore, take it that the approximations (21) and (24) which we have seen hold good to the second order when expansion in series is convergent, will probably not be numerically very far out when $\mathbf{M}_1$ has a large value, in which case the normal method of development cannot be used.

It is important, at this stage, and to justify the above assertion, to consider a few numerical examples.

Tables I. and II. give the values of $\Delta x_2$ and $\sin \alpha_2$ for a single refracting surface, calculated for a number of values of $\mathbf{M}_1$ and two inclinations in each case. The inclinations are fixed from the perpendicular distance $\varpi$ of $\mathrm{A}_1$ from the incident ray. This, for moderate inclinations, is sensibly the same as the intercept made by the incident ray on the principal plane. $\varpi$ has been given the two values 0·5 and 0·25 in every case, except for $\mathbf{M}_1 = 2$ where $\varpi = 0{\cdot}5$ leads to a physically impossible value. In this case $\varpi = 0{\cdot}25$ and $\varpi = 0{\cdot}125$ have been used to define the ray.

In each case four values have been computed (1) the correct one, from trigonometrical calculation; (2) the values given by formulæ (21) and (24)—these are shown in the column headed "fractional formula"; (3) the values obtained by expansion in series, up to the optician's first order of aberrations inclusive, that is including $\sin^3 \alpha_0$ in the development of $\Delta x_2$ and $\sin \alpha_2$—these are shown in the column headed "first order"; (4) the same series carried to the second order of aberrations inclusive, *i.e.*, to the terms involving $\sin^5 \alpha_0$—these are shown in the column headed "second order." It should be noted that these first and second order approximations are the most accurate that can be obtained, much more so than more usual ones, proceeding in powers of $\sin \alpha_2$ or $\tan \alpha_2$.

TABLE I.—Values of $\Delta x_2$ for Single Refracting Surface.

| $\mathbf{M}$. | $\varpi$. | First order. | Second order. | Fractional formula. | True. |
|---|---|---|---|---|---|
| 10 | 0·5 | − 33·35293 | − 107·6252 | + 27·18563 | + 26·66794 |
| 10 | 0·25 | − 8·33823 | − 12·98025 | − 18·81008 | − 18·85273 |
| 2 | 0·25 | − 0·250000 | − 0·378906 | − 0·516129 | − 0·527526 |
| 2 | 0·125 | − 0·062500 | − 0·070557 | − 0·071749 | − 0·071781 |
| 0·5 | 0·5 | − 0·015625 | − 0·016541 | − 0·016598 | − 0·016611 |
| 0·5 | 0·25 | − 0·003906 | − 0·003963 | − 0·003964 | − 0·003963 |
| 0 | 0·5 | − 0·166667 | − 0·174769 | − 0·175183 | − 0·175809 |
| 0 | 0·25 | − 0·041667 | − 0·042173 | − 0·042179 | − 0·042189 |
| − 1 | 0·5 | − 1·000000 | − 0·947500 | − 0·950119 | − 0·956680 |
| − 1 | 0·25 | − 0·250000 | − 0·246719 | − 0·246761 | − 0·246838 |

TABLE II.—Values of Sin $\alpha_2$ for Single Refracting Surface.

| $\mathbf{M}$. | $\varpi$. | First order. | Second order. | Fractional formula. | True. |
|---|---|---|---|---|---|
| 10 | 0·5 | + 0·0250865 | + 0·0454644 | + 0·0576348 | + 0·0610770 |
| 10 | 0·25 | − 0·0078936 | − 0·0072568 | − 0·0071911 | − 0·0071828 |
| 2 | 0·25 | − 0·2187500 | − 0·2065430 | − 0·1987180 | − 0·1978219 |
| 2 | 0·125 | − 0·1210938 | − 0·1207123 | − 0·1206710 | − 0·1206691 |
| 0·5 | 0·5 | 0·2539063 | 0·2541962 | 0·2542195 | 0·2542230 |
| 0·5 | 0·25 | 0·1254883 | 0·1254974 | 0·1254975 | 0·1254974 |
| 0 | 0·5 | 0·1805556 | 0·1817130 | 0·1826667 | 0·1827294 |
| 0 | 0·25 | 0·0850694 | 0·0851267 | 0·0851286 | 0·0851291 |
| − 1 | 0·5 | 0·1250000 | 0·1299375 | 0·1311526 | 0·1314354 |
| − 1 | 0·25 | 0·0531250 | 0·0532793 | 0·0532873 | 0·0532884 |

It appears from the above that the fractional formulæ are not merely equal, but appreciably superior to the second order formulæ, and this not merely in cases such as those of the three first entries in Table I., in which the convergency of the series for $\Delta x_2$ is either absent or slow, but in *every case* where the fractional or second order formulæ differ sensibly from the true value. (Clearly a divergence of 1 in the last place cannot be claimed as significant, for the last figure in Tables I. and II. is probably not correct within $\pm 2$, in some cases.) An estimate of the range of the formula can be obtained from the fact that in the cases, $\varpi = 0·5$, $\mathbf{M}_1 = 10$ and 2, the angles of incidence were 52° 34′ and 48° 35′ respectively, and, for the other values, angles of incidence of 20° and 30° are quite common.

In view of this the accuracy of the results is surprising and, from the point of view of the further applications of the method, most encouraging.

## § 7. *Combination of Two Systems.*

Call the systems 1 and 3, and the initial, intermediate and final media 0, 2, 4.

$f_1$, $f_3$ are the focal lengths of the systems, as defined in § 2. $M_1$, $\mathbf{M}_1$ are the transverse and ray magnifications in the first system, $M_3$, $\mathbf{M}_3$ in the second system.

$M_3+\Delta M_3$, $\mathbf{M}_3+\Delta\mathbf{M}_3$ refer to the transverse and ray magnifications in the second system when $I_2$, the true intersection of ray 2 with the axis, is taken as the object point for the second refraction (instead of $J_2$, which refers to transverse and ray magnifications $M_3$, $\mathbf{M}_3$).

Using the notation of §§ 2, 5, we assume

$$\Delta x_2 = n_2 f_1 (A_1 t_2^2 + E_1 t_2^4)/(1+B_1 t_2^2) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (41)$$

$$q_2 = t_2 (1+B_1 t_2^2)/(1+C_1 t_2^2) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (42)$$

where $q_2 = \begin{matrix}\sin\\ \tan\end{matrix}\{\alpha_2\}$, $t_2 = \begin{Bmatrix}\sin\gamma_2\\ \tan\beta_2\end{Bmatrix}$, and $B_1$, $C_1$ have suitable forms according as sines or tangents are considered. The constant $E_1$ is zero if the systems reduce to single refracting surfaces. Its form in the more general case will be discussed later.

If we denote by $\Delta_1 x_4$ that part of $\Delta x_4$ which is due to $\Delta x_2$ and by $\Delta_3 x_4$ that which is introduced by the aberrations proper to the system 3,

$$\Delta_1 x_4 = -n_4 f_3 \,\Delta M_3$$

where $\Delta M_3$ is obtained from $\Delta x_2$ by means of

$$\Delta x_2 = n_2 f_3 \{1/(M_3+\Delta M_3) - 1/M_3\},$$

leading to

$$\Delta M_3 = -M_3^2\,\Delta x_2/(n_2 f_3 + M_3\,\Delta x_2).$$

Thus

$$\Delta_1 x_4 = n_4 f_1 M_3^2 (A_1 t_2^2 + E_1 t_2^4)/\{1 + t_2^2 (B_1 + M_3 A_1 f_1/f_3)\} \quad . \quad . \quad . \quad . \quad . \quad (43)$$

Again

$$\Delta_3 x_4 = \frac{n_4 f_3 \{(A_3+\Delta A_3)\, q_2^2 (\mathbf{M}_3+\Delta\mathbf{M}_3)^{-2} + (E_3+\Delta E_3)\, q_2^4 (\mathbf{M}_3+\Delta\mathbf{M}_3)^{-4}\}}{1+(B_3+\Delta B_3)\, q_2^2 (\mathbf{M}_3+\Delta\mathbf{M}_3)^{-2}}$$

$$= n_4 f_3 [\{A_3\mathbf{M}_3^{-2} + \Delta M_3\, d\,(A_3\mathbf{M}_3^{-2})/dM_3\} q_2^2 + E_3 t_2^4 \mathbf{M}_3^{-4}]/(1+B_3 t_2^2 \mathbf{M}_3^{-2}), \quad (44)$$

retaining only terms of second order in $t_2^2$. Writing now for $q_2^2$ in the above its " first order " equivalent $t_2^2 + 2\,(B_1 - C_1)\, t_2^4$, we have

$$\Delta x_4/n_4 = \Delta_1 x_4/n_4 + \Delta_3 x_4/n_4$$

$$= \frac{\begin{array}{l}\{f_1 M_3^2 A_1 + f_3 A_3/\mathbf{M}_3^2\}\, t_2^2 + t_2^4\, [f_1 \{B_3 A_1 M_3^2 \mathbf{M}_3^{-2} + E_1 M_3^2 - A_1 M_3^2\, d\,(A_3\mathbf{M}_3^{-2})/dM_3 \\ + A_1 A_3 M_3 \mathbf{M}_3^{-2}\} + f_3 \{B_1 A_3 \mathbf{M}_3^{-2} + 2 A_3 (B_1 - C_1)\, \mathbf{M}_3^{-2} + E_3 \mathbf{M}_3^{-4}\}]\end{array}}{\{1 + t_2^2 (B_1 + A_1 M_3 f_1/f_3)\}\{1 + B_3 t_2^2/\mathbf{M}_3^2\}}, \quad . \quad . \quad . \quad . \quad (45)$$

Remembering that $t_2 = \mathbf{M}_3 t_4$ and retaining only the first two terms of the denominator product

$$\Delta x_4/n_4 = (f_{13}A_{13}t_4{}^2 + f_{13}E_{13}t_4{}^4)/(1+B_{13}t_4{}^2) \quad . \quad . \quad . \quad . \quad . \quad . \quad (46)$$

where

$$f_{13}A_{13} = f_3A_3 + f_1A_1M_3{}^2\mathbf{M}_3{}^2 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (47)$$

$$B_{13} = B_3 + B_1\mathbf{M}_3{}^2 + A_1M_3\mathbf{M}_3{}^2 f_1/f_3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (48)$$

$$\begin{aligned} f_{13}E_{13} = {} & f_3E_3 + f_1E_1M_3{}^2\mathbf{M}_3{}^4 \\ & + f_1A_1\mathbf{M}_3{}^2\{B_3M_3{}^2 + A_3M_3 - M_3{}^2\mathbf{M}_3{}^2\, d(A_3\mathbf{M}_3{}^{-2})/dM_3\} \\ & + f_3A_3\mathbf{M}_3{}^2\{3B'_1 - 2C_1\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (49) \end{aligned}$$

Again

$$q_4 = \frac{q_2(\mathbf{M}_3 + \Delta\mathbf{M}_3)^{-1}\{1 + (B_3 + \Delta B_3)\, q_2{}^2(\mathbf{M}_3 + \Delta\mathbf{M}_3)^{-2}\}}{1 + (C_3 + \Delta C_3)\, q_2{}^2(\mathbf{M}_3 + \Delta\mathbf{M}_3)^{-2}},$$

and retaining only terms of order $t_2{}^3$

$$\begin{aligned} q_4 &= q_2(\mathbf{M}_3 + \Delta\mathbf{M}_3)^{-1}(1 + B_3t_2{}^2\mathbf{M}_3{}^{-2})/(1 + C_3t_2{}^2\mathbf{M}_3{}^{-2}) \\ &= t_2\mathbf{M}_3{}^{-1}(1 - \Delta\mathbf{M}_3/\mathbf{M}_3)(1 + B_1t_2{}^2)(1 + B_3t_2{}^2\mathbf{M}_3{}^{-2})/\{(1 + C_1t_2{}^2)(1 + C_3t_2{}^2\mathbf{M}_3{}^{-2})\} \\ &= t_2\mathbf{M}_3{}^{-1}(1 + B_1t_2{}^2 + B_3t_2{}^2\mathbf{M}_3{}^{-2} + t_2{}^2A_1M_3f_1/f_3)/(1 + C_1t_2{}^2 + C_3t_2{}^2\mathbf{M}_3{}^{-2}) \\ &= t_4(1 + B_{13}t_4{}^2)/(1 + C_{13}t_4{}^2), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (50) \end{aligned}$$

where $B_{13}$ has the value given by (48) and

$$C_{13} = C_3 + C_1\mathbf{M}_3{}^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (51)$$

The equations (47), (48), (49), (51), give the constants for the combined system in terms of those for the components. It may appear at first sight as if the choice of the constants $B_{13}$ and $E_{13}$ had been arbitrary, for clearly, if $\lambda$ be any quantity,

$$\{f_{13}A_{13}t_4{}^2 + f_{13}(E_{13} + \lambda A_{13})t_4{}^4\}/\{1 + (B_{13} + \lambda)t_4{}^2\}$$

will give a development equally valid to the second order. But, if we do this, and we wish to preserve the simple character of the relation (48) giving the B for the combination, $\lambda$ will have to be a linear function of $A_1$, $A_3$, $B_1$, $B_3$. $\lambda A_{13}$ must then necessarily contain terms of one or other of the forms $A_1B_1$, $A_1{}^2$, $A_3B_3$, $A_3{}^2$. Thus the new E will contain such terms and will no longer be of type (49) which is linear in the aberration coefficients of each system taken separately. Thus the lineo-linear type of equation for $E_{13}$ requires $\lambda = 0$.

We note also that the equations of combination are identical in form, whether we are dealing with the sine or the tangent of the inclination as argument.

### § 8. *Nature of the Quantities* A, B, C, E *in the General Case of any System.*

In the case of the single refracting surface we found that A, B, C were polynomials of degrees 4, 3, 2 in $\mathbf{M}$, and that E was identically zero.

In addition, for such a surface, equations (23), (25) and (27) show that the coefficient of $\mathbf{M}_1^4$ in A is $n_0/n_2$ times the coefficient of $\mathbf{M}_1^3$ in **B** or B. This may be otherwise stated in the form :—

$$\mathrm{A}_1 - n_0\mathbf{M}_1\mathrm{B}_1/n_2, \textit{ i.e.}, \ \mathrm{A}_1 - \mathrm{M}_1\mathrm{B}_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\mathrm{I})$$

is of the fourth degree only in appearance and reduces to an expression of the third degree in $\mathbf{M}_1$ or $\mathrm{M}_1$.

The same holds good for $\mathrm{A}_1 - \mathrm{M}_1\mathbf{B}_1$, so this result is independent of whether the sine or tangent is taken as argument. The same remark applies to all the results of the present section and to the other invariant relations shortly to be proved. We may therefore conveniently state it here once for all.

If we now refer to the equations (47), (48), and remember that in any combination—

$$\mathrm{M}_{13}/f_3 = \mathrm{M}_3/f_{13} - 1/f_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (52)$$

and

$$1/f_1\mathrm{M}_{13} = 1/f_{13}\mathrm{M}_1 - 1/f_3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (53)$$

with the corresponding equations

$$n_0\mathbf{M}_{13}/f_3 = n_2\mathbf{M}_3/f_{13} - n_4/f_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

$$n_4/f_1\,\mathbf{M}_{13} = n_2/f_{13}\,\mathbf{M}_1 - n_0/f_3 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (55)$$

and the obvious conditions

$$\mathbf{M}_{13} = \mathbf{M}_1\mathbf{M}_3\,; \qquad \mathrm{M}_{13} = \mathrm{M}_1\mathrm{M}_3,$$

we note first that, if $\mathrm{A}_3$ is a quartic in $\mathbf{M}_3$ it is also a quartic in $\mathrm{M}_{13}$ or $\mathbf{M}_{13}$, and that if $\mathrm{A}_1$ is a quartic in $\mathbf{M}_1$, it becomes, on multiplication by $\mathrm{M}_3^2\mathbf{M}_3^2$, a quartic in $\mathrm{M}_{13}$ or $\mathbf{M}_{13}$, since $\mathrm{M}_3 = n_2\mathbf{M}_3/n_4$, and therefore $\mathbf{M}_1^r\mathrm{M}_3^2\mathbf{M}_3^2 = n_2^2\mathbf{M}_{13}^r\mathbf{M}_3^{4-r}/n_4^2$, which makes every term in $\mathrm{A}_1\mathrm{M}_3^2\mathbf{M}_3^2$ a quartic in $\mathbf{M}_{13}$, because $\mathbf{M}_3$ is a linear function of $\mathbf{M}_{13}$ and $4-r$ is here zero or positive.

(47) then shows that $\mathrm{A}_{13}$ will be a quartic function of $\mathbf{M}_{13}$, if $\mathrm{A}_1$ and $\mathrm{A}_3$ are quartic functions of $\mathbf{M}_1$ and $\mathbf{M}_3$ respectively. But we know this to be the case for a single refracting surface. Hence it holds good of any system compounded of such surfaces.

Now consider (48). If $\mathrm{B}_3$ is a cubic in $\mathbf{M}_3$ it becomes a cubic in $\mathbf{M}_{13}$.

Again, if $\mathrm{A}_1 - \mathrm{M}_1\mathrm{B}_1 =$ a cubic $\mathrm{U}_1$ in $\mathbf{M}_1$

$$\begin{aligned}\mathbf{M}_3^2(\mathrm{B}_1 + \mathrm{A}_1\mathrm{M}_3f_1/f_3) &= \mathbf{M}_3^2(\mathrm{B}_1\{1 + \mathrm{M}_{13}f_1/f_3\} + \mathrm{U}_1\mathrm{M}_3f_1/f_3) \\ &= \mathbf{M}_3^2\mathrm{M}_3f_1(\mathrm{B}_1/f_{13} + \mathrm{U}_1/f_3),\end{aligned}$$

using (52), and by reasoning similar to the one given for $A_{13}$ the last expression is a cubic in $\mathbf{M}_{13}$.

Hence $B_{13}$ is a cubic in $\mathbf{M}_{13}$.

Consider now $A_{13}-M_{13}B_{13}$. This is found, after some reductions, and using (52), to be

$$(f_3/f_{13})\,(A_3-M_3B_3)+f_3B_3/f_1+M_3\mathbf{M}_3^2\,(A_1-M_1B_1).$$

Of the above terms, $A_3-M_3B_3$ is a cubic in $\mathbf{M}_3$ and therefore also a cubic in $\mathbf{M}_{13}$. $B_3$ is a cubic in $\mathbf{M}_{13}$. $A_1-M_1B_1$ is a cubic in $\mathbf{M}_1$ and when multiplied by $M_3\mathbf{M}_3^2$ becomes a cubic in $\mathbf{M}_{13}$.

Hence if the condition (I) holds good for the components, it also holds good for the resultant system. But we have seen that it holds for a single refracting system; thus it holds for any combination. Also B will be a cubic in M for any system.

As regards C, examination of (51), remembering that for a single surface $C_3$ and $C_1$ are quadratics in $\mathbf{M}_3$, $\mathbf{M}_1$ respectively, leads immediately to the conclusion that C is a quadratic in $\mathbf{M}$ for any system.

We now come to the coefficient E. Here the single refracting surface gives no precedent for $E_1$ and $E_3$. Let us examine the other terms in $E_{13}$. These can be written in the form $f_1A_1\mathbf{M}_3^2M_3^2\,(B_3-\,dA_3/dM_3)+3A_3\mathbf{M}_3^2\,(f_3B_1+f_1M_3A_1)-2f_3A_3\mathbf{M}_3^2C_1$, and, using $A_1-B_1M_1 = U_1$; $A_3-B_3M_3 = U_3$, where $U_1$, $U_3$ are then cubics in $M_1$, $M_3$ respectively, this is found to reduce to

$$-f_1A_1\mathbf{M}_3^2M_3^2\,(M_3\,dB_3/dM_3+dU_3/dM_3)+3A_3\mathbf{M}_3^2M_3f_1\,(U_1+f_3B_1/f_{13})-2f_3A_3\mathbf{M}_3^2C_1. \quad (56)$$

Now

$$A_1\mathbf{M}_3^2M_3^2 = \text{quartic in } M_{13}.$$

$$M_3\,dB_3/dM_3+dU_3/dM_3 = \text{cubic in } M_3 = \text{cubic in } M_{13}.$$

$$\mathbf{M}_3^2M_3\,(U_1+f_3\mathbf{B}_1/f_{13}) = \text{cubic in } M_3 = \text{cubic in } M_{13}.$$

$$\mathbf{A}_3 = \text{quartic in } M_{13}.$$

$$\mathbf{M}_3^2C_1 = \text{quadratic in } M_{13}.$$

Hence the three terms in (56) are of form

(quartic) (cubic) + (quartic) (cubic) + (quartic) (quadratic),

and this leads to a rational integral polynomial of degree 7 in $M_{13}$.

Further consideration, however, shows that it is of degree 7 only in appearance, for the terms which can lead to expressions of degree 7 in $M_{13}$ are clearly

$$-f_1\mathbf{M}_3^2M_3^2A_1\,dB_3/dM_3+3A_3f_1\mathbf{M}_3^2M_3\,(U_1+f_3B_1/f_{13}),$$

or, dropping the factor $f_1$

$$\begin{aligned}\mathbf{M}_3{}^2\mathrm{M}_3\{3\mathrm{A}_3(\mathrm{U}_1+f_3\mathrm{B}_1/f_{13})-\mathrm{M}_3{}^2\mathrm{A}_1 d\mathrm{B}_3/d\mathrm{M}_3\}\\ =\mathbf{M}_3{}^2\mathrm{M}_3\{3\mathrm{B}_3\mathrm{M}_3(\mathrm{U}_1+f_3\mathrm{B}_1/f_{13})+3\mathrm{U}_3(\mathrm{U}_1+f_3\mathrm{B}_1/f_{13})\\ -\mathrm{M}_3{}^2\mathrm{M}_1\mathrm{B}_1\, d\mathrm{B}_3/d\mathrm{M}_3-\mathrm{M}_3{}^2\mathrm{U}_1\, d\mathrm{B}_3/d\mathrm{M}_3\}.\end{aligned}$$

The term

$$3\mathrm{U}_3\mathbf{M}_3{}^2\mathrm{M}_3(\mathrm{U}_1+f_3\mathrm{B}_1/f_{13})$$

is clearly of the form cubic × cubic and the terms leading to expressions of 7th degree reduce to

$$\mathbf{M}_3{}^2\mathrm{M}_3\{\mathrm{M}_3\mathrm{U}_1(3\mathrm{B}_3-\mathrm{M}_3\, d\mathrm{B}_3/d\mathrm{M}_3)+\mathrm{M}_3\mathrm{B}_1(3\mathrm{B}_3 f_3/f_{13}-\mathrm{M}_{13}\, d\mathrm{B}_3/d\mathrm{M}_3)\},$$

and since $d\mathrm{M}_{13}/d\mathrm{M}_3=f_3/f_{13}$, this can be written

$$\mathbf{M}_3{}^2\mathrm{M}_3\{\mathrm{M}_3\mathrm{U}_1(3\mathrm{B}_3-\mathrm{M}_3\, d\mathrm{B}_3/d\mathrm{M}_3)+(\mathrm{M}_3\mathrm{B}_1 f_3/f_{13})(3\mathrm{B}_3-\mathrm{M}_{13}\, d\mathrm{B}_3/d\mathrm{M}_{13})\}.$$

But since $\mathrm{B}_3$ is a cubic, in either $\mathrm{M}_3$ or $\mathrm{M}_{13}$, $3\mathrm{B}_3-\mathrm{M}_3\, d\mathrm{B}_3/d\mathrm{M}_3$ and $3\mathrm{B}_3-\mathrm{M}_{13}\, d\mathrm{B}_3/d\mathrm{M}_{13}$ are both quadratics in $\mathrm{M}_3$ or $\mathrm{M}_{13}$.

The above expression therefore reduces to $\mathrm{M}_3$ × sum of two quantities each of form: cubic in $\mathrm{M}_{13}$ × quadratic in $\mathrm{M}_{13}$, that is, to a sextic in $\mathrm{M}_{13}$.

We see, therefore, that those terms in $\mathrm{E}_{13}$ which do not involve $\mathrm{E}_1$ or $\mathrm{E}_3$ are a polynomial of sixth degree in $\mathrm{M}_{13}$ or $\mathbf{M}_{13}$.

It follows that for a *lens* E is necessarily a sextic in the magnification.

Suppose now that $\mathrm{E}_1$ and $\mathrm{E}_3$ are both sextics in $\mathbf{M}_1$, $\mathbf{M}_3$ respectively. Then $\mathrm{E}_3$ is a sextic in $\mathrm{M}_{13}$ and $\mathrm{E}_1\mathbf{M}_3{}^4\mathrm{M}_3{}^2$ will also be a sextic in $\mathrm{M}_{13}$, that is $\mathrm{E}_{13}$ will again be a sextic in $\mathbf{M}_{13}$.

Hence, since any system is built up of combinations of lenses or single refracting surfaces, we find that E is a sextic polynomial in $\mathbf{M}$ for any system.

Examination of particular cases shows that E is not, in general, divisible by A, so that the vanishing of the latter does not usually involve the disappearance of the second order terms.

### §9. *Invariant Relations.*

Certain relations exist between the coefficients A, B, C, E which remain the same in form, whatever the number of refracting surfaces. One of these we have already dealt with, namely the fact that

$$\mathrm{A}-\mathrm{MB}$$

reduces to an expression of the third degree, *i.e.*, the coefficients of highest degree in M in A and B are the same.

This we shall refer to as the *first* invariant relation (I.).

A second invariant relation takes the form

$$\mathrm{B}-\mathrm{C}=\tfrac{3}{8}(1-\mathbf{M}^2)+\tfrac{1}{4}\, d\mathrm{A}/d\mathrm{M}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad \text{(II)}$$

when we use $\tan\beta_{2n}$ as argument, and

$$\mathrm{B}-\mathrm{C} = -\tfrac{1}{8}(1-\mathbf{M}^2)+\tfrac{1}{4}\,d\mathrm{A}/d\mathrm{M}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (\mathrm{II}')$$

when we use $\sin\gamma_{2n}$ as argument.

That these relations hold good for the single refracting surface is readily verified from equations (22), (23), (25), (27) and (32). Suppose now that, for systems 1 and 3 separately, the relation

$$\mathrm{B}-\mathrm{C} = \sigma\,(1-\mathbf{M}^2)+\tfrac{1}{4}\,d\mathrm{A}/d\mathrm{M}$$

holds, where $\sigma = \frac{3}{8}$ or $-\frac{1}{8}$ according to the nature of the argument, then from (48) and (51)

$$\begin{aligned}\mathrm{B}_{13}-\mathrm{C}_{13} &= \mathrm{B}_3-\mathrm{C}_3+\mathbf{M}_3^{\,2}\,(\mathrm{B}_1-\mathrm{C}_1)+\mathrm{A}_1\mathrm{M}_3\mathbf{M}_3^{\,2}f_1/f_3\\ &= \sigma\,(1-\mathbf{M}_{13}^{\,2})+\tfrac{1}{4}\,d\mathrm{A}_3/d\mathrm{M}_3+\tfrac{1}{4}\mathbf{M}_3^{\,2}\,(d\mathrm{A}_1/d\mathrm{M}_1+4\mathrm{A}_1\mathrm{M}_3f_1/f_3), \quad . \quad . \qquad (57)\end{aligned}$$

and from (47)

$$f_{13}\,d\mathrm{A}_{13}/d\mathrm{M}_{13} = f_3\,d\mathrm{A}_3/d\mathrm{M}_{13}+f_1\{\mathrm{M}_3^{\,2}\mathbf{M}_3^{\,2}\,d\mathrm{A}_1/d\mathrm{M}_{13}+\mathrm{A}_1\,d\,(\mathrm{M}_3^{\,2}\mathbf{M}_3^{\,2})/d\mathrm{M}_{13}\}.$$

But

$$d\mathrm{M}_{13} = (f_3/f_{13})\,d\mathrm{M}_3 = (f_1/f_{13})\,\mathrm{M}_3^{\,2}\,d\mathrm{M}_1.$$

Hence

$$d\mathrm{A}_{13}/d\mathrm{M}_{13} = d\mathrm{A}_3/d\mathrm{M}_3+\mathbf{M}_3^{\,2}\,d\mathrm{A}_1/d\mathrm{M}_1+(f_1\mathrm{A}_1/f_3)\,d\,(\mathrm{M}_3^{\,2}\mathbf{M}_3^{\,2})/d\mathrm{M}_3,$$

and since $\mathbf{M}_3/\mathrm{M}_3 = \text{const.}$, the last differential coefficient is $4\mathrm{M}_3\mathbf{M}_3^{\,2}$.

Using this result (57) becomes

$$\mathrm{B}_{13}-\mathrm{C}_{13} = \sigma\,(1-\mathbf{M}_{13}^{\,2})+\tfrac{1}{4}\,d\mathrm{A}_{13}/d\mathrm{M}_{13},$$

which is of the same form as the equation we started from. Hence, if the two components of the compound system satisfy the second invariant relation, the resultant system also satisfies it. But we have seen that the relation holds good for single refracting surfaces—hence it holds good universally.

It should be noted that the second invariant relation is really a *first order* relation and connects the first order aberration of the inclination of a ray, with the first order longitudinal spherical aberration.

### § 10. *The Constants* A, B, C, E *for an Optical System Reversed and for Negative Lenses.*

Certain important general relations are found to hold between the constants A, B, C, E for rays going through an optical system and the corresponding constants A′, B′, C′, E′, for the same system reversed, and by making use of them we can obtain either set from the other.

We arrive most simply at these relations as follows:—If after traversing the system we retrace our steps, the result is equivalent to compounding the system with itself reversed, with the difference that, in the second set of refractions, the measurement of length parallel to the axis is reversed in direction. An examination of the equations (41) *et seq.*, § 7, on which the formulæ of combination are based, shows that this is analytically equivalent to changes in the sign of the focal length in the second set of refractions.

We have therefore

$$f_1 = f, \quad f_3 = -f, \quad \mathrm{M}_1 = \mathrm{M}, \quad \mathrm{M}_3 = 1/\mathrm{M}, \quad \mathbf{M}_1 = \mathbf{M}, \quad \mathbf{M}_3 = 1/\mathbf{M}, \quad \mathbf{M}_{13} = \mathrm{M}_{13} = 1,$$

and we also find that $f_{13} = \infty$. But $f_{13}\mathrm{A}_{13}$, and $f_{13}\mathrm{E}_{13}$ have definite limiting values, and as $f_{13}$ does not otherwise explicitly enter into the equations of combination, no difficulty arises on that account.

Now, after retracing our steps in this way, we necessarily arrive at a perfect image, so that $\Delta x_4 \equiv 0$ and $\tan \alpha_4 = \tan \beta_4$, leading to

$$f_{13}\mathrm{A}_{13} \equiv 0,$$

$$f_{13}\mathrm{E}_{13} \equiv 0,$$

and

$$\mathrm{B}_{13} - \mathrm{C}_{13} \equiv 0.$$

These lead to the following identical relations

$$\mathrm{A}(\mathrm{M})/\mathrm{M}^2\mathbf{M}^2 - \mathrm{A}'(\mathrm{M}^{-1}) \equiv 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (58)$$

$$\begin{aligned}\mathrm{E}(\mathrm{M})/\mathrm{M}^2\mathbf{M}^4 - \mathrm{E}'(\mathrm{M}^{-1}) + \mathrm{A}(\mathrm{M})\,\mathbf{M}^{-2}\{\mathrm{B}'(\mathrm{M}^{-1})\,\mathrm{M}^{-2} + 3\mathrm{A}'(\mathrm{M}^{-1})\,\mathrm{M}^{-1} + d\mathrm{A}'(\mathrm{M}^{-1})/d\mathrm{M}\} \\ -\mathrm{A}'(\mathrm{M}^{-1})\,\mathbf{M}^{-2}\{3\mathrm{B}(\mathrm{M}) - 2\mathrm{C}(\mathrm{M})\} \equiv 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (59)\end{aligned}$$

$$\mathrm{B}'(\mathrm{M}^{-1}) - \mathrm{C}'(\mathrm{M}^{-1}) + \mathbf{M}^{-2}\{\mathrm{B}(\mathrm{M}) - \mathrm{C}(\mathrm{M})\} - \mathrm{A}(\mathrm{M})\,\mathrm{M}^{-1}\mathbf{M}^{-2} \equiv 0. \quad . \quad . \quad (60)$$

Equation (58) may be written in either of the two forms

$$\mathrm{A}(\mathrm{M}) \equiv \mathrm{M}^2\mathbf{M}^2\mathrm{A}'(\mathrm{M}^{-1})$$

$$\mathrm{A}(\mathbf{M})/n_0^2 \equiv \mathbf{M}^4\mathrm{A}'(\mathbf{M}^{-1})/n_2^2. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\mathrm{III})$$

This we shall refer to as the third invariant relation. It shows that, if we divide A by the square of the initial refractive index, the coefficients of powers of $\mathbf{M}$ equidistant from the beginning and end of the development are interchanged by reversing the system. Equation (59) becomes on multiplying up by $\mathrm{M}^2\mathbf{M}^4$, using (58) and simplifying

$$\mathrm{E} - \mathrm{M}^2\mathbf{M}^4\mathrm{E}' + \mathrm{A}\{\mathbf{M}^2\mathrm{B}' - \mathrm{A}/\mathrm{M} - 3\mathrm{B} + 2\mathrm{C} + d\mathrm{A}/d\mathrm{M}\} \equiv 0, \quad . \quad . \quad . \quad (61)$$

omitting the arguments M, 1/M of A, B, B′, &c., since no confusion can occur.

Use now the second invariant relation

$$d\mathrm{A}/d\mathrm{M} \equiv 4\mathrm{B} - 4\mathrm{C} - 4\sigma(1 - \mathbf{M}^2),$$

(61) becomes

$$\mathrm{E}-\mathrm{M}^2\mathbf{M}^4\mathrm{E}'+\mathrm{A}\left\{\mathrm{B}+\mathbf{M}^2\mathrm{B}'-2\mathrm{C}-\mathrm{A}/\mathrm{M}-4\sigma\left(1-\mathbf{M}^2\right)\right\}\equiv 0. \quad . \quad . \quad . \quad (62)$$

Now substitute from (60) for $\mathrm{B}+\mathbf{M}^2\mathrm{B}'$ the value

$$\mathrm{C}+\mathbf{M}^2\mathrm{C}'+\mathrm{A}/\mathrm{M},$$

and (61) leads to

$$\mathrm{E}-\mathrm{M}^2\mathbf{M}^4\mathrm{E}'+\mathrm{A}\left\{\mathbf{M}^2\mathrm{C}'-\mathrm{C}-4\sigma\left(1-\mathbf{M}^2\right)\right\}\equiv 0 \quad . \quad . \quad . \quad . \quad . \quad (63)$$

For a single refracting surface, where E, E′ are identically zero, this must lead to

$$\mathbf{M}^2\mathrm{C}'-\mathrm{C} = 4\sigma\left(1-\mathbf{M}^2\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\mathrm{IV})$$

a result which is easily verified from equation (32).

Now consider a system compounded of two systems. For the system direct, we have

$$\mathrm{C}_{13} = \mathrm{C}_3+\mathbf{M}_3{}^2\mathrm{C}_1.$$

Similarly, for the system reversed, change $\mathrm{C}_3$ into $\mathrm{C}'_1$, $\mathrm{C}_1$ into $\mathrm{C}'_3$, $\mathbf{M}_3$ into $1/\mathbf{M}_1$.

$$\begin{aligned}\mathrm{C}_{31} &= \mathrm{C}'_1+\mathbf{M}_1{}^{-2}\mathrm{C}'_3\\ \mathbf{M}_{13}{}^2\mathrm{C}_{31}-\mathrm{C}_{13} &= \mathbf{M}_1{}^2\mathbf{M}_3{}^2\mathrm{C}'_1-\mathbf{M}_3{}^2\mathrm{C}_1+\mathbf{M}_3{}^2\mathrm{C}'_3-\mathrm{C}_3\\ &= \mathbf{M}_3{}^2\left(\mathbf{M}_1{}^2\mathrm{C}'_1-\mathrm{C}_1\right)+\mathbf{M}_3{}^2\mathrm{C}'_3-\mathrm{C}_3\end{aligned}$$

and using (IV) which we know to be true for a single refracting surface

$$\begin{aligned}\mathbf{M}_{13}{}^2\mathrm{C}_{31}-\mathrm{C}_{13} &= 4\sigma\left[\mathbf{M}_3{}^2\left(1-\mathbf{M}_1{}^2\right)+1-\mathbf{M}_3{}^2\right]\\ &= 4\sigma\left(1-\mathbf{M}_{13}{}^2\right).\end{aligned}$$

In other words (IV) will hold for the resultant system if it holds for the components, and therefore as in previous similar cases, it holds for any system.

We shall call (IV) the *fourth* invariant relation.

Equation (63) then shows that there exists a ***fifth*** invariant relation

$$\mathrm{E} = \mathrm{M}^2\mathbf{M}^4\mathrm{E}'. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (\mathrm{V})$$

or

$$\mathrm{E}\left(\mathbf{M}\right)/n_0{}^2 = \mathbf{M}^6\mathrm{E}'\left(\mathbf{M}^{-1}\right)/n_2{}^2,$$

so that E possesses a property similar to that of A, previously noticed, viz., if we divide it by the square of the initial refractive index, the coefficients of powers of $\mathbf{M}$ equidistant from the beginning and end of the development in $\mathbf{M}$ are interchanged.

Equation (59) has therefore led us to *two* independent invariant relations.

On the other hand it will be found that (60) leads to no new relation. For if we substitute into it for $\mathrm{B}'-\mathrm{C}'$ and $\mathrm{B}-\mathrm{C}$ in virtue of the second invariant relation, and then use the third relation, it becomes an identity.

It gives, however, on using (IV) to eliminate $C'$,

$$\mathbf{M}^2B'+B = A/M+2C+4\sigma(1-\mathbf{M}^2) \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (64)$$

which is a convenient form for calculating $B'$.

If now A, B, C, E are known for any system, the corresponding quantities are immediately obtainable for the reversed system, $A'$, $B'$, $C'$, $E'$ being given by equations (III), (64), (IV) and (V) respectively.

This will generally halve the labour of calculations, if it is found desirable to tabulate these constants for a complete set of lenses. It will then be sufficient to start from the equi-convex lens and vary the curvatures in one sense only.

Incidentally we note also that the second invariant relation enables us to find C, so soon as A and B are known, so that only A, B and E require to be calculated.

The aberration constants for a reversed system have a further important application in the case of lenses. Consider a positive lens (fig. 4), the initial ray converging to $I_0$ and the final ray to $I_4$. If now we interchange the full and dotted portions of the initial and final rays in fig. 4, we obtain, since here the initial and final media are the same, the case of a ray going through a lens in which the front and back character of the two surfaces have been interchanged. In fact $r_1$ and $r_3$ have been interchanged and the sign of the thickness $c_2$ has been reversed. This leads to a negative lens, of the same *numerical* power as the original positive lens, and with the same mean curvature, but a negative thickness. Such a lens, of course, is not physically realisable, although a part of it can be physically obtained by rotating the wedge *beyond* the intersection V of the two surfaces.

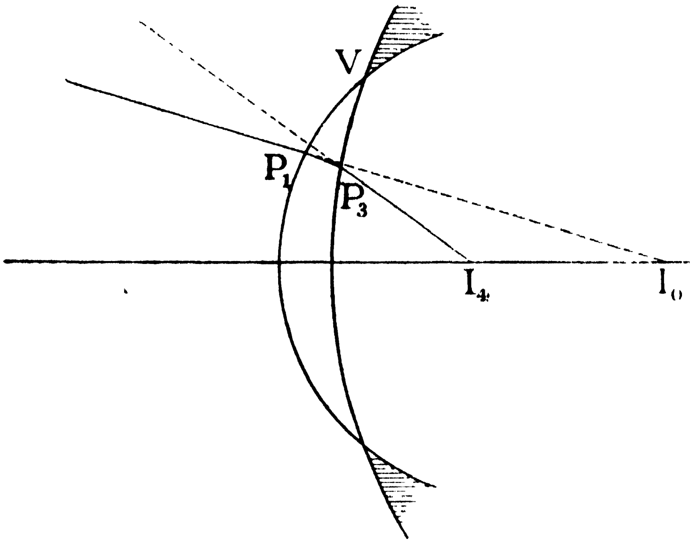


Fig. 4.

But, in the case of the ideally *thin* lenses, where the thickness is zero, the ideally thin positive and negative lenses, having the same mean curvature and numerical power, correspond in this way.

Now $I_4$ and $I_0$ are also interchanged. If we consider $I_4$ as the initial point, then we are really considering a set of rays starting from $I_4$ in the last medium and travelling backwards through the original positive lens. In other words, the aberration constants for the corresponding negative lens are identical with those for the original positive lens reversed, and the equations (III), (64), (IV) and (V) are applicable to calculate them.

This, again, will greatly diminish the work of calculation. In the case of ideally thin negative lenses, we see that A, B, C, E are directly obtained from the corresponding thin positive lenses. In the case of *thick* negative lenses the corresponding positive lens has a negative thickness.

Now for various reasons it will probably be convenient, in calculating A, B, C, E for lenses, to express them in the form

$$A_c = A_0 + c\,(dA/dc)_0,$$

&c., where $c$, the thickness, is small, as it usually is in practice, and $A_0$ refers to an ideally thin lens.

When the formulæ are put in this form, it is perfectly simple to calculate $A_{-c}$, $B_{-c}$, &c., and then to obtain the corresponding results for the negative lens with a positive thickness.

### § 11. *Explicit Values of* A, B, C, E *for a Thick Lens* (*Tangent Formula*).

For the purposes of numerical calculation and comparison with correct trigonometrically found values, we have worked out explicitly the form of the expressions A, B, C, E for a thick lens, when we use $\tan\beta_4$ as the argument; the formulæ are expressed in terms of the focal lengths of each surface and of the combination and the thickness does not appear explicitly. The initial and final media being the same $n_0 = n_4$ and we have written $n = n_2/n_0$.

The work of algebraic calculation has been straightforward but extremely heavy, and we therefore omit it here entirely, the object being to publish the results for reference, in case other workers desire to use them for tabulation purposes, but it is hardly to be expected that designing opticians should work direct from the algebraic expressions as they stand.

For the purpose of this section we shall write

$$A = A_0 + A_1M + A_2M^2 + A_3M^3 + A_4M^4.$$

$$B = B_0 + B_1M + B_2M^2 + B_3M^3.$$

$$C = C_0 + C_1M + C_2M^2.$$

$$E = E_0 + E_1M + E_2M^2 + E_3M^3 + E_4M^4 + E_5M^5 + E_6M^6.$$

The suffixes here have a different meaning to that which has been previously ascribed to them, but no confusion is likely to arise on this account.

The values are as follows, $f$ denoting the focal length of the lens, $f_1$, $f_3$ the focal lengths of the surfaces, so that

$$f_1 = r_1/(n-1), \qquad f_3 = -r_3/(n-1)$$

$$A_0 = \frac{1}{2n^2(n-1)^2}\left[-\frac{f_1+f_3}{f_1^4}f^3 + \frac{(n+1)^2 f_3 f^2}{f_1^3} - 2n(1+n+n^2)\frac{f_3 f}{f_1^2} + n^2(n+1)^2\frac{f_3}{f_1} - \frac{n^4 f_3}{f}\right|$$

$$A_1 = \frac{1}{2n^2(n-1)^2}\left[-\frac{4(f_1+f_3)}{f_1^3 f_3}f^3 + 4(n+1)^2\frac{f^2}{f_1^2} - 4n(1+n+n^2)\frac{f}{f_1} + n^2(n+1)^2\right|$$

$$A_2 = \frac{1}{2n^2(n-1)^2}\left[-\frac{6(f_1+f_3)}{f_1^2 f_3^2}f^3 + 6(n+1)^2\frac{f^2}{f_1 f_3} + 2n(1+n+n^2)f\left(\frac{1}{f_1}+\frac{1}{f_3}\right)\right|$$

$A_3 = A_1$ with $f_1$ and $f_3$ interchanged

$A_4 = A_0$ ,, ,, ,,

$$B_0 = \frac{1}{4n^2(n-1)^2}\left[-\frac{2f^3}{f_1^2}\left(\frac{1}{f_1}+\frac{1}{f_3}\right) + 2(n^2-n+1)\frac{f^2}{f_1^2} + n(n-1)^2\frac{f}{f_1} - n^2(n^2-n+1)\right|$$

$$B_1 = \frac{1}{4n^2(n-1)^2}\left[-\frac{6f^3}{f_1 f_3}\left(\frac{1}{f_1}+\frac{1}{f_3}\right) + 6(n^2+1)\frac{f^2}{f_1 f_3} + n(n-1)^2 f\left(\frac{1}{f_1}+\frac{1}{f_3}\right)\right]$$

$$B_2 = \frac{1}{4n^2(n-1)^2}\left[-\frac{6f^3}{f_3^2}\left(\frac{1}{f_1}+\frac{1}{f_3}\right) + 6(n^2+n+1)\frac{f^2}{f_3^2} - 3n(n+1)^2\frac{f}{f_3} + 3n^2(n^2-n+1)\right|$$

$B_3 = A_4$

$C_0 = B_0 - \frac{3}{8} - \frac{1}{4}A_1$

$C_1 = B_1 - \frac{1}{2}A_2$

$C_2 = B_2 + \frac{3}{8} - \frac{3}{4}A_3$

$$E_0 = \frac{3}{8(n-1)^4}\left[\begin{aligned} &-\frac{(n^2-n+1)}{n^4}\frac{(f_1+f_3)f^5}{f_1^6} + \frac{(n+1)^2}{n^4}f^4\left\{\frac{n}{f_1^4} + \frac{(n^2-n+1)f_3}{f_1^5}\right\} \\ &-\frac{f^3}{n^3}\left\{\frac{n^3+3n^2+n}{f_1^3} + \frac{2(n^4+n^2+1)f_3}{f_1^4}\right\} + \frac{f^2}{n^2}\left\{\frac{2n^2}{f_1^2} + (n+1)^2(n^2-n+1)\frac{f_3}{f_1^3}\right\} \\ &\qquad -f(n^2-n+1)\frac{f_3}{f_1^2} \end{aligned}\right]$$

$$E_1 = \frac{3}{8(n-1)^4}\left[\begin{aligned} &-\frac{6(n^2-n+1)}{n^4}\frac{(f_1+f_3)f^5}{f_3 f_1^5} + \frac{(n+1)^2}{n^4}f^4\left\{\frac{5n}{f_3 f_1^3} + \frac{6(n^2-n+1)}{f_1^4} + \frac{n f_3}{f_1^5}\right\} \\ &-\frac{f^3}{n^3}\left\{\frac{4(n^3+3n^2+n)}{f_3 f_1^2} + \frac{(n+1)^4+8(n^4+n^2+1)}{f_1^3} + (n+1)^4\frac{f_3}{f_1^4}\right\} \\ &+\frac{f^2}{n^2}\left\{\frac{6n^2}{f_3 f_1} + \frac{4(n+1)^2(n^2+1)}{f_1^2} + \frac{2(n-1)^2(n^2+n+1)f_3}{f_1^3}\right\} \\ &-\frac{f}{n}\left\{\frac{2n(2n^2+n+2)}{f_1} + \frac{(n+1)^4 f_3}{f_1^2}\right\} + \frac{n f_3}{f_1}(n+1)^2 \end{aligned}\right]$$

$$
\begin{aligned}
E_2 = \frac{3}{8(n-1)^4}\Bigg[ & -\frac{15(n^2-n+1)}{n^4}\frac{f^5(f_1+f_3)}{f_1^4 f_3^2}+\frac{(n+1)^2}{n^4}f^4\left\{\frac{10n}{f_1^2 f_3^2}+\frac{15(n^2-n+1)}{f_1^3 f_3}+\frac{5n}{f_1^4}\right\} \\
& -\frac{f^3}{n^3}\Bigg\{\frac{6(n^3+3n^2+n)}{f_1 f_3^2}+\frac{12(n^4+n^2+1)+4(n+1)^4}{f_1^2 f_3} \\
& \qquad +\frac{4(n+1)^4+2(n^4+n^2+1)}{f_1^3}+\frac{(n^3+3n^2+n)f_3}{f_1^4}\Bigg\} \\
& +\frac{f^2}{n^2}\Bigg\{\frac{6n^2}{f_3^2}+\frac{6(n+1)^2(n^2+n+1)}{f_1 f_3}+\frac{8(n+1)^2(n^2+n+1)}{f_1^2} \\
& \qquad +\frac{(n+1)^2(n^2+3n+1)f_3}{f_1^3}\Bigg\} \\
& -\frac{f}{n}\Bigg\{\frac{4n(n+1)^2+n^3-n^2+n}{f_3}+\frac{2(n+1)^4+2(n^2+n+1)(n^2+3n+1)}{f_1} \\
& \qquad +\frac{2(n^2+n+1)(n^2+3n+1)f_3}{f_1^2}\Bigg\} \\
& +n\{(n+1)^2+4(n^2+n+1)\}+(n+1)^2(n^2+3n+1)\frac{f_3}{f_1} \\
& \qquad -n^2(n^2+3n+1)\frac{f_3}{f_1}\Bigg]
\end{aligned}
$$

$$
\begin{aligned}
E_3 = \frac{3}{8(n-1)^4}\Bigg[ & -\frac{20(n^2-n+1)}{n^4}\frac{f^5(f_1+f_3)}{f_1^3 f_3^3}+\frac{(n+1)^2}{n^4}f^4\Bigg\{\frac{10n}{f_3 f_1}\left(\frac{1}{f_1^2}+\frac{1}{f_3^2}\right) \\
& \qquad +\frac{20(n^2-n+1)}{f_1^2 f_3^2}\Bigg\} \\
& -\frac{f^3}{n^3}\left\{4(n^3+3n^2+n)\left(\frac{1}{f_1^3}+\frac{1}{f_3^3}\right)+\frac{8(n^4+n^2+1)+6(n+1)^4}{f_1 f_3}\left(\frac{1}{f_1}+\frac{1}{f_3}\right)\right\} \\
& +\frac{f^2}{n^2}\left\{2n^3\left(\frac{f_1}{f_3^3}+\frac{f_3}{f_1^3}\right)+4(n+1)^4\left(\frac{1}{f_1^2}+\frac{1}{f_3^2}\right)+\frac{12(n+1)^2(n^2+n+1)}{f_1 f_3}\right\} \\
& -\frac{f}{n}\Bigg\{2n(n+1)^2\left(\frac{f_1}{f_3^2}+\frac{f_3}{f_1^2}\right)+[(n+1)^4 \\
& \qquad +4(n^2+n+1)(n^2+3n+1)]\left(\frac{1}{f_1}+\frac{1}{f_3}\right)\Bigg\} \\
& +4n(n^2+n+1)\left(\frac{f_1}{f_3}+\frac{f_3}{f_1}\right)+2(n+1)^2(n^2+3n+1) \\
& -2n^2(n+1)^2\frac{(f_1+f_3)}{f}+2n^4\frac{f_1 f_3}{f^2}\Bigg]
\end{aligned}
$$

$E_4 = E_2$ with $f_1$ and $f_3$ interchanged.

$E_5 = E_1$ ,, ,, ,, ,,

$E_6 = E_0$ ,, ,, ,, ,,

It should be noted carefully that all the above refer to expressions in terms of the tangent of the Gaussian inclination, this being the argument we have used in the numerical work.

### § 12. *Values of* A, B, C, E *for a Thin Lens.*

When the lens is thin, we have the relation

$$\frac{1}{f_1}+\frac{1}{f_3}=\frac{1}{f},$$

which enables the values of § 11 to be considerably simplified.

In this case it is useful to introduce a quantity K such that

$$\mathrm{K}=\frac{\text{mean curvature of the lens}}{\text{power of the lens}}=\frac{f}{2}\left(\frac{1}{r_1}+\frac{1}{r_3}\right).$$

When this is done the constants A, B, C, E take the following forms :—

$$\begin{aligned}\mathrm{A} = &-\{(1-\mathrm{M})^2/2n\}\,\{(n+2)\,[(1-\mathrm{M})\,\mathrm{K}-(1+\mathrm{M})\,(n+1)/(n+2)]^2\\ &\qquad +n^3\,(1-\mathrm{M})^2/4\,(n-1)^2-n^2\,(1+\mathrm{M})^2/4\,(n+2)\}.\end{aligned}$$

$$\begin{aligned}\mathrm{B} = {}&(1-\mathrm{M})^2\,\mathrm{K}^2\,\{n-1-(n+2)\,\mathrm{M}\}/2n+(1-\mathrm{M})\,\mathrm{K}\,(1+\mathrm{M}+4\mathrm{M}^2)\,(n+1)/4n\\ &+(1-\mathrm{M})\,\{\mathrm{M}\,(1+\mathrm{M})/4n+(1-\mathrm{M})\,[3n-2n^2-3+\mathrm{M}\,(6n-4n^2-3)]/8\,(n-1)^2\}.\end{aligned}$$

$$\mathrm{C} = -3\,(1-\mathrm{M})^2\,\mathrm{K}^2/2n+3\,(1-\mathrm{M}^2)\,\mathrm{K}\,(n+1)/4n-\tfrac{3}{4}\,(1-\mathrm{M}^2)-\tfrac{3}{8}n\,(1-\mathrm{M})^2/(n-1)^2.$$

$$\mathrm{E} = \frac{3\,(1-\mathrm{M})^2}{128\,(n-1)^4\,n^3}\left[\begin{aligned}&(1+\mathrm{M}^4)\,(-4n^5+8n^4-n^3-4n^2+3n-1)\\ &\quad+\mathrm{M}\,(1+\mathrm{M}^2)\,(8n^6-16n^5+4n^4+4n^3-12n^2+12n-4)\\ &\quad+\mathrm{M}^2\,(-16n^7+32n^6-8n^5-8n^4+10n^3-16n^2+18n-6)\\ &\quad+8\,(n-1)\,(1-\mathrm{M})\,(1+\mathrm{M})\,\mathrm{K}\,\{(1+\mathrm{M})^2\\ &\qquad\qquad(-2n^5+5n^4-2n^3-3n^2+3n-1)\\ &\qquad\qquad+\mathrm{M}\,(2n^6-8n^4+4n^3+2n^2)\}\\ &\quad+8\,(n-1)^2\,(1-\mathrm{M})^2\,\mathrm{K}^2\,\{(1+\mathrm{M})^2\\ &\qquad\qquad(-2n^5+8n^4-7n^3-6n^2+9n-3)\\ &\qquad\qquad-2\mathrm{M}\,n^2\,(n^2-2n-1)\}\\ &\quad+16\,(n-1)^3\,(1+\mathrm{M})\,(1-\mathrm{M})^3\,\mathrm{K}^3\,\{2n^4-4n^3-2n^2+6n-2\}\\ &\quad+16\,(n-1)^4\,(1-\mathrm{M})^4\,\mathrm{K}^4\,(-n^3+3n-1)\end{aligned}\right]$$

We notice that when M = 1 (which gives one of the zeros of A) B, C and E all vanish with it, and also E/A remains finite. Hence in this case, the term $\mathrm{E}t_4{}^4$ will not rise in importance, even when A = 0. But in this case A may have two other real zeros, and these are not zeros of E, so that E plays an important part in the neighbourhood of such zeros.

## § 13. *Numerical Test for a Single Lens.*

To test the formulæ, a number of longitudinal aberrations were calculated trigonometrically for five positive lenses of refractive index 1·52, unit focal length and thickness $\frac{1}{16}$. The first was a meniscus-shaped lens for which $r_1 = 0·349418$, $r_3 = 1$.

The second was a plano-convex lens, of which the convex side is towards the incoming light. The third was an equi-convex lens. The fourth and fifth were the second and first reversed. If K has the meaning defined in § 12, the values of K for these five lenses are 1·93094, 0·96154, 0, −0·96154 and − 1·93094 respectively, so that they proceed by approximately equal steps of K.

The constants A, B, E were calculated for these five lenses and the longitudinal aberrations computed from the formula. The rays selected met the first principal plane at a distance 0·15 from the axis, corresponding to an aperture *f*/3·4, nearly.

The results are shown in Table III.

Table III.—Longitudinal Aberrations of Five Selected Lenses.

| M. | Lens 1. | | | Lens 2. | | | Lens 3. | | |
|---|---|---|---|---|---|---|---|---|---|
| | Formula. | True. | Percentage error. | Formula. | True. | Percentage error. | Formula. | True. | Percentage error. |
| 3 | -7·98683 | -7·87422 | 1·41 | -1·12450 | -1·12814 | 0·32 | -0·32378 | -0·32439 | 0·19 |
| 2 | -1·55287 | -1·52334 | 1·94 | -0·40167 | -0·40283 | 0·29 | -0·14240 | -0·14282 | 0·29 |
| 0·5 | -0·003527 | -0·003530 | 0·09 | -0·011697 | -0·011705 | 0·07 | -0·033008 | -0·033096 | 0·27 |
| 0 | -0·06028 | -0·06043 | 0·24 | -0·024728 | -0·024749 | 0·08 | -0·036194 | -0·036213 | 0·05 |
| -0·5 | -0·20902 | -0·20916 | 0·21 | -0·083863 | -0·083950 | 0·10 | -0·057067 | -0·057101 | 0·06 |
| -1 | -0·42692 | -0·42772 | 0·19 | -0·18428 | -0·18448 | 0·11 | -0·094736 | -0·094787 | 0·05 |
| -2 | -1·00383 | 1·00531 | 0·15 | -0·48874 | -0·48930 | 0·12 | -0·21707 | -0·21724 | 0·08 |

| M. | Lens 4. | | | Lens 5. | | |
|---|---|---|---|---|---|---|
| | Formula. | True. | Percentage error. | Formula. | True. | Percentage error. |
| 3 | -0·098969 | -0·099086 | 0·12 | -0·10401 | -0·10444 | 0·41 |
| 2 | -0·046089 | -0·046128 | 0·30 | -0·013671 | -0·013686 | 0·11 |
| 0·5 | -0·073959 | -0·074535 | 0·77 | -0·14762 | -0·15068 | 2·03 |
| 0 | -0·099843 | -0·100044 | 0·20 | -0·22172 | -0·22324 | 0·68 |
| -0·5 | -0·13967 | -0·13984 | 0·12 | -0·32828 | -0·32966 | 0·42 |
| -1 | -0·19084 | -0·19101 | 0·09 | -0·45753 | -0·45887 | 0·29 |
| -2 | -0·32454 | -0·32476 | 0·07 | -0·77417 | -0·77511 | 0·12 |

The mean percentage error of these results is 0·34, so that the formula determines, on the average, the longitudinal aberration correct to 1 part in 300. The bulk of this, however, is contributed by three cases, namely: M = 3 and 2 for lens 1 where the aberrations are very large and differ very widely from the usual first and second order approximations, so that, although the error of the formula approaches 2 per cent., it nevertheless represents a great improvement upon these approximations; and M = 0·5 for lens 5, which corresponds to extreme curvature and highest inclination, so that one of the angles of refraction is as great as 48½ degrees. Even here the table below shows that the formula is an appreciable improvement on the usual second-order approximation. If these three cases are omitted, the mean percentage error works out to be about 0·21, so that in general the formula determines the longitudinal aberration correct to about 1 part in 500.

It is interesting to note what the usual first and second order approximations lead to in a few cases.

| | First order approximation. | Percentage error. | Second order approximation. | Percentage error. |
|---|---|---|---|---|
| Lens 1, M = 3 . . . | -1·34012 | 83·0 | -2·45132 | 68·8 |
| „ 1, M = 2 . . . | -0·48545 | 68·1 | -0·82104 | 46·1 |
| „ 1, M = -2 . . . | -1·348142 | 34·1 | -0·88456 | 12·0 |
| „ 2, M = 3 . . . | -0·67356 | 40·3 | -0·94436 | 16·3 |
| „ 5, M = 0·5 . . . | -0·12136 | 19·5 | -0·14307 | 5·1 |

This gives a measure of the numerical improvement effected by the fractional formula whenever the usual method of approximation is seriously out, even though in none of the cases above does the convergency of the series actually fail.

In the above the series are in powers of tan $\beta_2$. Had they been taken in powers of tan $\alpha_2$, as is frequently done, the first and second order approximations would have been far worse.

One interesting outcome of these calculations relates to the relative importance of the terms in $Et_4^4$ and $At_4^2$. The ratio $Et_4^2/A$ is small in every case taken (of course these exclude the neighbourhood of points where A = 0, where naturally E becomes of great importance). But for the set of magnifications taken, the greatest ratio of the second term to the first is less than 0·03 and the mean value of this ratio is only 0·0082, so that, in fact, the E term—although so complicated algebraically—does not exercise any great influence numerically.

This is important, as it shows that, at any rate for lenses, it does not require to be computed with anything like the same order of accuracy which is needed for A and B.

## § 14. *The Singular Inclination and Convergency Factor for any System.*

Referring again to fig. 2 we see that $\lambda = \alpha_2$ and $F_0I_0 = -\Delta x_2$ for rays proceeding through the system reversed and initially parallel.

Thus using accents, as before, to denote the coefficients and inclinations for the system reversed, and noting that the accented coefficients all refer to zero magnification, we have, using *tangents*

$$\tan\lambda = \tan\beta'_2\,(1+B'_0\tan^2\beta'_2)/(1+C'_0\tan^2\beta'_2)$$

and

$$-F_0I_0 = fn_0\,(A'_0\tan^2\beta'_2+E'_0\tan^4\beta'_2)/(1+B'_0\tan^2\beta'_2).$$

Here the suffixes in the A, B, C, &c., A′, B′, C′, &c., have the same meaning as in § 11.

Thus

$$-1/M = (A'_0\tan^2\beta'_2+E'_0\tan^4\beta'_2)/(1+B'_0\tan^2\beta'_2),$$

whence, developing $\cot^2\beta'_2$ in descending powers of M and stopping at the second term

$$\cot^2\beta'_2 = -A'_0M-B'_0+E'_0/A'_0.$$

Substituting into

$$\cot^2\lambda = \cot^2\beta'_2+2\,(C'_0-B'_0)$$

which is valid to the same order of approximation, we obtain

$$\cot^2\lambda = -A'_0M+2\,(C'_0-B'_0)-B'_0+E'_0/A'_0 \quad . \quad . \quad . \quad . \quad . \quad (67)$$

as the second approximation for the singular inclination when M is large, the first approximation being $\cot^2\lambda = -A'_0M$.

To the same order the convergency factor is

$$1+\tan^2\alpha_0\,(A'_0M+\{2\,(B'_0-C'_0)+B'_0-(E'_0/A'_0)\}),$$

*i.e.*,

$$1+(n_2^2\tan^2\beta_2/n_0^2)\,\{A'_0M^3+(3B'_0-2C'_0-E'_0/A'_0)\,M^2\}, \quad . \quad . \quad . \quad . \quad (68)$$

$\beta_2$ referring to a ray passing through the system in the standard sense.

Now, using the equations (III), (64), (IV) and (V) of § 10 and equating suitable coefficients, we find that

$$\begin{aligned} A'_0 &= (n_0^2/n_2^2)\,A_4, & E'_0 &= (n_0^4/n_2^4)\,E_6,\\ C'_0 &= (n_0^2/n_2^2)\,C_2-4\sigma, & B'_0 &= (n_0^2/n_2^2)\,(A_3+2C_2-B_2)-4\sigma\\ 3A_3 &= 4B_2-4C_2+4\,(n_2^2/n_0^2)\,\sigma, & A_4 &= B_3, \end{aligned}$$

whence, after substitution, (68) becomes

$$1+\tan^2\beta_2\,(B_3M^3+\{B_2-E_6/A_4\}\,M^2). \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (69)$$

Now, if our B leads to a sound approximation to the convergency factor for M large, this should be

$$1+B\tan^2\beta_2,$$

or, to the same approximation which we have been using

$$1+\tan^2\beta_2\,(\,B_3M^3+B_2M^2)\,. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (70)$$

We see, therefore, that the development of the correct convergency factor in descending powers of M will give a result which always agrees with our B, so far as the highest term in M is concerned, but makes the term in $M^2$ in general different.

In the case of a lens $E_6/A_4$ is in general small, compared with $B_2$, so that this discrepancy makes little difference, but it may well be that, when we come to deal with more complicated systems, this will not be the case.

A little consideration, however, shows that, when this is so, our formula is very readily corrected so as to take this difficulty into account, without involving any lengthy numerical computation.

If we consider the formula

$$\Delta x_2 = n_2 f\,\{At_2^2+(E-AE_6M^2/A_4)\,t_2^4\}/\{1+(B-E_6M^2/A_4)\,t_2^2\}$$

it is clear that it leaves the development of $\Delta x_2$ in powers of $t_2$ unaltered as far as the second order inclusive. It alters the coefficient $B_2$ of B so as to make the two leading terms agree with (69). It also alters E in such a way as to remove the term in $M^6$ and reduce E to a quintic. In fact it gives for the new E the remainder obtained after the first step in the division of E by A, according to the usual process.

In practice, the terms in $(E_6/A_4)\,M^2$ are very readily added as follows:

$$\Delta x_2 = n_2 f\,\frac{At_2^2+Et_2^4/(1-E_6M^2t_2^2/A_4)}{1+Bt_2^2/(1-E_6M^2t_2^2/A_4)}$$

and this amounts to applying the same corrective factor $1/(1-E_6M^2t_2^2/A_4)$ or $1/(1-En_0^2t_0^2/n_2^2A_4)$ to the second terms in both numerator and denominator. This factor, expressed in terms of the inclination of the incident ray, is independent of the magnification, and a short table will enable it to be found in any given case without difficulty.

A similar correction has then to be made in C; in order to keep the development of $\tan\alpha_2$ the same we must have

$$\tan\alpha_2 = t_2\,\{1+(B-E_6M^2/A_4)\,t_2^2\}\,/\,\{1+(C-E_6M^2/A_4)\,t_2^2\},$$

and writing this as

$$\frac{t_2\,\{1+Bt_2^2/(1-E_6M^2t_2^2/A_4)\}}{1+Ct_2^2/(1-E_6M^2t_2^2/A_4)}$$

we see that the same corrective factor has to be applied to all the second terms in the formulæ.

In the above we have used $\tan\beta_2$ as our argument, but the formulæ and the correction take precisely the same form if $\sin\gamma_2$ is the argument.

If we apply this corrective factor to the first two entries of Table III., which give a large percentage error—these correspond to cases approaching the failure of convergency and are therefore critical, we find, for lens (1)

| M. | True aberration. | Formula uncorrected. | Percentage error. | Formula corrected. | Percentage error. |
|---|---|---|---|---|---|
| 3 | −7·87422 | −7·98683 | 1·41 | −7·90811 | 0·43 |
| 2 | −1·52334 | −1·55287 | 1·94 | −1·54089 | 1·16 |

which shows a very sensible improvement.

The significance of this alteration is brought out more clearly when we consider the limiting case $M = \infty$ that is, rays actually issuing from the front focus (the case of an eye-piece). In this case, the geometrical image being at infinity, it is inconvenient to define the emergent ray by means of either longitudinal or transverse aberrations.

Let us consider the intercept of the ray on the back focal plane.

$$\begin{aligned}\text{This} &= (-n_2 f\mathrm{M} + \Delta x_2)\tan\alpha_2\\ &= n_2 f\,|\,(\mathrm{A}\tan^2\beta_2 + e\tan^4\beta_2)/(1 + b\tan^2\beta_2) - \mathrm{M}\,|\tan\alpha_2.\end{aligned}$$

Also

$$\tan\alpha_2 = (\tan\alpha_0/\mathbf{M})\,(1 + b\mathbf{M}^{-2}\tan^2\alpha_0)/(1 + c\mathbf{M}^{-2}\tan^2\alpha_0),$$

where

$$b = \mathrm{B} - \mathrm{E}_6\mathrm{M}^2/\mathrm{A}_4,\quad c = \mathrm{C} - \mathrm{E}_6\mathrm{M}^2/\mathrm{A}_4,\quad e = \mathrm{E} - \mathrm{E}_6\mathrm{A}\mathrm{M}^2/\mathrm{A}_4,$$

and it is clear that, in the limit, where M (and $\mathbf{M}$) $= \infty$, $\tan\alpha_2$ must be *finite.*

This requires that $b$ and $c$ shall be of order $\mathrm{M}^3$ and $\mathrm{M}^2$ respectively, which is right, and leads to

$$\tan\alpha_2 = nb_3\tan^3\alpha_0/n\,(1 + c_2\tan^2\alpha_0),$$

$b_3$ and $c_2$ being the coefficients of $\mathrm{M}^3$ and $\mathrm{M}^2$ in $b$ and $c$ respectively.

On the other hand, it is equally obvious that the intercept on the back focal plane must also approach a definite limit. Hence the factor

$$(\mathrm{A}\tan^2\beta_2 + e\tan^4\beta_2)/(1 + b\tan^2\beta_2) - \mathrm{M},$$

*i.e.*,

$$\frac{-1 + (\mathrm{A} - \mathrm{M}b)\tan^2\beta_2/\mathrm{M} + e\tan^4\beta_2/\mathrm{M}}{1/\mathrm{M} + b\tan^2\beta_2/\mathrm{M}},$$

or

$$\frac{-1 + (\mathrm{A} - \mathrm{M}b)\tan^2\alpha_0/\mathrm{M}\mathbf{M}^2 + e\tan^4\alpha_0/\mathrm{M}\mathbf{M}^4}{1/\mathrm{M} + b\tan^2\alpha_0/\mathrm{M}\mathbf{M}^2}$$

must tend to a finite limit as M approaches $\infty$.

This necessarily involves that (1) $\mathrm{A} - \mathrm{M}b$ has $\mathrm{M}^3$ in its leading term—a result already established, and (2) that $e$ involves $\mathrm{M}^5$ (and not $\mathrm{M}^6$) in its leading term.

Hence, whenever we deal with incident rays actually passing through the front focus—and this necessarily occurs as soon as the results of the present paper are applied to aberrations off the axis—the *modified* B, C and E have to be used.

### § 15. *Combination Formulæ for More than Two Systems.*

The combination formulæ (47), (48), (51) are capable of explicit generalisation for any number of systems.

In the case of A and C successive applications of equations (47) and (51) lead at once to the results

$$\begin{aligned} f_{135\,..\,2n+1}\mathrm{A}_{135\,..\,2n+1} = {} & (f_1\mathrm{A}_1)\,\mathrm{M}^2{}_{35\,...\,2n+1}\mathbf{M}^2{}_{35\,...\,2n+1} \\ & +(f_3\mathrm{A}_3)\,\mathrm{M}^2{}_{5\,..\,2n+1}\mathbf{M}^2{}_{5\,...\,2n+1}+\,\ldots \\ & +(f_{2n-1}\mathrm{A}_{2n-1})\,\mathrm{M}^2{}_{2n+1}\mathbf{M}^2{}_{2n+1}+f_{2n+1}\mathrm{A}_{2n+1}. \quad . \qquad (71) \end{aligned}$$

$$\begin{aligned} \mathrm{C}_{135\,...\,2n+1} = {} & \mathrm{C}_1\mathbf{M}^2{}_{35\,..\,2n+1}+\mathrm{C}_3\mathbf{M}^2{}_{5\,...\,2n+1}+\,\ldots \\ & +\mathrm{C}_{2n-1}\mathbf{M}^2{}_{2n+1}+\mathrm{C}_{2n+1}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (72) \end{aligned}$$

in which we have reverted to the notation of §§ 2, 7.

In the case of B, we have for three systems

$$\begin{aligned} \mathrm{B}_{135} & = \mathrm{B}_{35}+\mathrm{B}_1\mathbf{M}_{35}{}^2+\mathrm{A}_1\mathrm{M}_{35}\mathbf{M}_{35}{}^2 f_1/f_{35} \\ & = \mathrm{B}_5+\mathrm{B}_3\mathbf{M}_5{}^2+\mathrm{B}_1\mathbf{M}_{35}{}^2+(f_1\mathrm{A}_1)\,\mathrm{M}_{35}\mathbf{M}_{35}{}^2/f_{35}+(f_3\mathrm{A}_3)\,\mathrm{M}_5\mathbf{M}_5{}^2/f_5, \end{aligned}$$

and the general law of combination is at once obvious.

We have

$$\begin{aligned} \mathrm{B}_{135\,...\,2n+1} = {} & \mathrm{B}_1\mathbf{M}^2{}_{35\,...\,2n+1}+\mathrm{B}_3\mathbf{M}^2{}_{5\,...\,2n+1}+\,\ldots\,+\mathrm{B}_{2n-1}\mathbf{M}^2{}_{2n+1}+\mathrm{B}_{2n+1} \\ & +(f_1\mathrm{A}_1)\,(\mathrm{M}\mathbf{M}^2/f\,)_{35\,..\,2n+1}+(f_3\mathrm{A}_3)\,(\mathrm{M}\mathbf{M}^2/f\,)_{5\,..\,2n+1}+\,\ldots\,+(f_{2n-1}\mathrm{A}_{2n-1})\,(\mathrm{M}\mathbf{M}^2/f\,)_{2n+1}. \quad (73) \end{aligned}$$

The case of E is more difficult and we have not been able to obtain a form in which it can be written down for the combination of *n* systems.

But we can deal with it as follows, by determining the contribution of any one component to the whole :—

Consider three systems 1, 3, 5. It will be convenient to look upon 3 as a single lens, forming part of a larger system. 1 will then be the system of lenses preceding 3, and 5 the system following.

Applying equation (49), simplified by the use of the second invariant relation, we find

$$\begin{aligned} f_{135}\mathrm{E}_{135} = {} & f_5\mathrm{E}_5+f_{13}\mathrm{E}_{13}\mathrm{M}_5{}^2\mathbf{M}_5{}^4 \\ & +f_{13}\mathrm{A}_{13}\mathbf{M}_5{}^2\mathrm{M}_5{}^2\,(3\mathrm{A}_5/\mathrm{M}_5-3\mathrm{B}_5+4\mathrm{C}_5+4\sigma\,\{1-\mathbf{M}_5{}^2\}) \\ & +3\mathrm{B}_{13}f_5\mathrm{A}_5\mathbf{M}_5{}^2-2\mathrm{C}_{13}f_5\mathrm{A}_5\mathbf{M}_5{}^2. \end{aligned}$$

Applying the equations of combination a second time, and picking out the terms involving $A_3$, $B_3$, $C_3$, $E_3$, we find these to be

$$\begin{aligned} & f_3E_3M_5{}^2\mathbf{M}_5{}^4+M_5{}^2\mathbf{M}_5{}^4\left[f_1A_1\mathbf{M}_3{}^2M_3{}^2\left(3A_3/M_3-3B_3+4C_3\right)+f_3A_3\mathbf{M}_3{}^2\left(3B_1-2C_1\right)\right] \\ & \quad +f_3A_3\mathbf{M}_5{}^2M_5{}^2\left[3A_5/M_5-3B_5+4C_5+4\sigma\left(1-\mathbf{M}_5{}^2\right)\right] \\ & \quad +3B_3f_5A_5\mathbf{M}_5{}^2-2C_3f_5A_5\mathbf{M}_5{}^2. \end{aligned}$$

Hence the lens 3 contributes to the final E

$$\begin{aligned} & f_3A_3\{3M_5\mathbf{M}_5{}^2\left[A_5+(f_1/f_3)\,A_1M_{35}\mathbf{M}_{35}{}^2\right]+4\sigma\left(1-\mathbf{M}_5{}^2\right)\mathbf{M}_5{}^2M_5{}^2 \\ & \qquad\qquad +M_5{}^2\mathbf{M}_5{}^2\left[-3B_5+4C_5+\left(3B_1-2C_1\right)\mathbf{M}^2_{35}\right]\} \\ & \quad +B_3\mathbf{M}_5{}^2\left(3f_5A_5-3f_1A_1\mathbf{M}_{35}{}^2M_{35}{}^2\right) \\ & \quad +C_3\mathbf{M}_5{}^2\left(-2f_5A_5+4f_1A_1\mathbf{M}_{35}{}^2M_{35}{}^2\right) \\ & \quad +f_3E_3M_5{}^2\mathbf{M}_5{}^4, \end{aligned}$$

and the A's, B's and C's in the curled brackets can be expressed in terms of the individual lenses of the system by means of equations (71), (72), (73).

If we denote the coefficients of $A_3$, $B_3$, $C_3$, $E_3$, in the above by $l_3$, $m_3$, $p_3$, $q_3$, then the contribution of the individual lens to E

$$= l_3A_3+m_3B_3+p_3C_3+q_3E_3.$$

Hence, if we vary $K_3$ for this lens, keeping focal length and magnifications unaltered

$$\Delta E = \Delta K_3\left(l_3\,\partial A_3/\partial K_3+m_3\,\partial B_3/\partial K_3+p_3\,\partial C_3/\partial K_3+q_3\,\partial E_3/\partial K_3\right).$$

If all the lenses are simultaneously varied, then we have

$$\Delta E = \Sigma\,\Delta K\left(l\,\partial A/\partial K+m\,\partial B/\partial K+p\,\partial C/\partial K+q\,\partial E/\partial K\right).$$

We have similar equations for $\Delta A$, $\Delta B$, $\Delta C$, but they take a much simpler form.

Using these, we can, if we have enough lenses, vary the K's so that, *between limits*, we can make our four constants A, B, C, E take up any assigned values, or, if we wish to keep any one constant whilst slightly varying the others, we have a linear relation between the $\Delta K$'s.

## § 16. *Conclusion.*

We have now established a formula of fractional type for the longitudinal aberration of a symmetrical system which, while algebraically correct as far as the second order, does in fact, give results beyond this order in those numerical cases which have been tried, and largely overcomes the difficulties of slow convergency in critical regions.

We have further obtained a method for calculating the coefficients of this formula for any symmetrical optical system in terms of the coefficients for the components, in such a way that the effect of any single component upon the whole combination is immediately obtained.

In considering the convergency of the series usually employed, we have found that the value of the approximation depends upon the particular variable employed, and that if we wish to avoid trouble owing to lack of convergency we must use sin $\alpha_0$ or tan $\alpha_0$ (or a suitable multiple of these) as argument, where $\alpha_0$ is the inclination of the original incident ray.

The numerical success of the new formula appears to suggest that progress in the algebraic treatment of symmetrical instruments is to be sought, not so much along the lines of developments in series, but in other mathematical directions such as continued products, or possibly continued fractions.

The next step would be to develop the method so as to cover the second order approximations to the emergent inclination. This will enable us to deal with aberrations off the axis of the system.

Some progress has already been made by the authors in this direction and the results will, it is hoped, form the subject of a later communication.

# III. *Investigations on Lightning Discharges and on the Electric Field of Thunderstorms.*

*By* C. T. R. Wilson, *F.R.S., Reader in Electrical Meteorology in the University of Cambridge and Observer in Meteorological Physics at the Solar Physics Observatory, Cambridge.*

(*Communicated by the Director of the Solar Physics Observatory.*)

Received March 15,—Read May 6, 1920.

[Plates 2–5.]

## Contents.

Section . . . Page
I. Methods of measurement . . . 74
II. Some typical records . . . 77
III. On the prevailing sign of the sudden changes produced in the potential gradient by lightning flashes . . . 85
IV. Magnitude of the changes produced in the electric field by lightning discharges at different distances . . . 86
V. Effects to be expected from different kinds of discharges at different distances . . . 87
VI. Electric moments of the discharges . . . 90
VII. Quantity of electricity discharged in an average lightning discharge . . . 91
VIII. Electric field of a thunder-cloud . . . 94
IX. Conditions determining discharge . . . 98
X. Dimensions of the regions discharged by lightning flashes . . . 101
XI. Maximum potential attained before the passage of a lightning discharge . . . 102
XII. Mean density of the charge in a thunder-cloud immediately before a discharge . . . 103
XIII. Charge associated with 1 c.c. of water in a cloud . . . 103
XIV. Disruption of drops by the electric field . . . 104
XV. Pressure within a charged portion of a cloud . . . 106
XVI. Thunder resulting from sudden contraction due to loss of charge . . . 106
XVII. Energy dissipated in lightning flashes . . . 106
XVIII. Interpretation of "recovery" curves . . . 107
XIX. Electrical currents maintained in the atmosphere by thunder-clouds and shower-clouds . 109
XX. Differences between the electrical effects of shower-clouds of positive and negative polarity . . . 110
XXI. The normal potential gradient and air-earth current of fine weather . . . 112
XXII. Influence of the nature of the earth's surface below a thunder-cloud or shower-cloud . . 113
XXIII. Secondary thunder-clouds . . . 114

[Published August 27, 1920.

THE observations discussed in this paper were made at the Solar Physics Observatory, Cambridge, mainly during the summer months of 1917.

## I. *Methods of Measurement.*

The method and apparatus used in the measurements are substantially those described in a paper "On Some Determinations of the Sign and Magnitude of Electric Discharges in Lightning Flashes."* The induced charge on an exposed earthed conductor (test-plate or sphere) is used as a measure of the electric field. The test-plate virtually forms part of a flat portion of the earth's surface, and the vertical electric force or potential gradient at ground level is equal (in electrostatic measure) to $4\pi Q/A$, where Q is the charge on its exposed surface and A is its area. The charge Q on the earth-connected sphere of radius R, when exposed at a height $h$, great compared with R, is a measure of the potential at that height; the zero potential of the sphere being the resultant of the undisturbed atmospheric potential V at the height $h$ and of the potential Q/R due to the charge on the sphere, so that $Q/R = -V$. The earthed conductors can be shielded from the earth's field: the test-plate by means of an earth-connected cover, the sphere by lowering it into a conducting case resting on the ground. The quantity of electricity which flows to earth through the connecting wire on exposing or shielding the test-plate or sphere, is measured by a special type of capillary electrometer in which the readings indicate the total quantity of electricity which has traversed the instrument; the sign and magnitude of the charge on the exposed conductor, and thus of the potential gradient, at the beginning and end of an exposure are thus determined. The sign and magnitude of sudden changes of potential gradient which occur while the conductor is exposed are indicated by the direction and magnitude of the resulting displacements of the electrometer meniscus. The total flow of electricity between the atmosphere and the test-plate or sphere during an exposure is also measured —being given by the difference between the electrometer readings before and after the exposure. The principal improvement introduced has been the provision of apparatus for giving a photographic trace of the electrometer readings; rapid changes in the field occupying less than one-tenth of a second are in this way recorded.

In the observations described in the previous paper the sphere was supported in a manner which did not admit of absolute measurements being made, as the charge measured included that on the upper part of the support as well as that on the sphere itself; in these earlier measurements therefore the sphere was standardised by comparison with the test-plate. The method of supporting the sphere is now such that the charge on the sphere alone is measured, while the disturbing effect of the earthed supporting rod is small, and thus the potential at the level of the earth-connected sphere can be calculated from the charge upon it. The new method of mounting the sphere is shown in fig. 1.

---

* 'Roy. Soc. Proc.,' A, vol. 92, p. 555, 1916.

The sphere, 30 cm. in diameter, is supported on an earth-connected brass tube B, 2 cm. in diameter, from which it is insulated by sulphur-coated ebonite E; insulators are indicated in the figure by the dotted areas. The tube is inserted within a wider one C which extends from the top to the bottom of the sphere and which is open below. The supporting tube B is rigidly fixed in a hole bored through the screw cap which closes the upper end of an iron pipe P, 5 cm. in external diameter and 427 cm. long, which can be turned about its lower end from the vertical to a nearly horizontal position as described in the former paper. The length of the brass tube from the top of the iron pipe to the bottom of the sphere is 38 cm. The connection between the sphere and the electrometer is made by means of a tightly stretched wire W supported by quartz insulators. The wire is not attached directly to the sphere but to a brass disc D insulated from the supporting tube and fitting loosely within the wider tube C inside the sphere. The sphere is fixed to the disc by means of a screw which projects from its inner surface and can thus readily be removed to give access to the insulation.

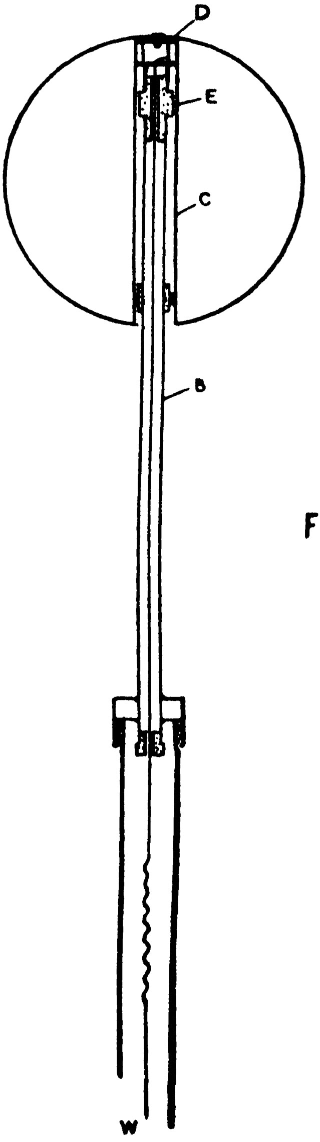


FIG I

When the sphere is exposed by raising the iron pipe to its vertical position the height of its centre above the ground is 480 cm.

The sphere when lowered is received in a metal-lined earthed box resting on the ground; a tightly fitting cover, also metal lined and earthed, protects the sphere from the atmospheric electrical field and from the weather. The charge on the earthed sphere in this position is taken as zero.

The charge $Q$ on an earthed sphere of radius $R$ at a height $h$ above level ground is assumed to be such that $Q/R-Q/2h+V = 0$, where $V$ is the undisturbed air potential at the height $h$. The presence of the neighbouring hut exerts a disturbing influence which however is not large: the correction to be applied has been estimated by imagining the hut to be replaced by a conducting hemisphere large enough to enclose it. The vertical potential gradient over level ground being assumed uniform throughout a height exceeding that of the hemisphere, the lowering of potential at a given point by the induced charge on the hemisphere is readily seen to be equal to $Va^3/r^3$ where $a$ is the radius of the hemisphere and $r$ is the distance of its centre from the given point. In the actual case the correction amounts to 6 per cent.

The charge on the exposed earthed test-plate (the surface of which is at ground level) is similarly diminished by the presence of the hut; the correction to be applied amounts in this case to about 1 per cent. A somewhat larger correction—estimated at 1·5 per cent.—has to be made for the effect of the induced charge of the earth-connected cover and its supporting arm. Apart from these small corrections the relation between the potential gradient F at ground level and the charge Q on the exposed earthed test-plate, of area A, is given by $4\pi Q/A = F$, when the quantities are expressed in C.G.S. electrostatic measure. The effective area of the plate is 2220 sq. cm.

For the measurement of the quantities of electricity which passed between the exposed conductor and earth through the connecting wire, the capillary electrometer described in the previous paper was used. By means of a $\frac{1}{4}$-inch microscope objective, placed with its axis vertical above the electrometer, an image of the meniscus was formed on a horizontal slit. The slit coincided in position with the image of the axis of the capillary tube and was almost in contact with the sensitive surface of a photographic plate kept in uniform motion at right angles to the slit. It was made by ruling a line with a razor blade on an exposed and developed "process" plate; it was protected by a strip of microscope cover-glass cemented with Canada balsam to the gelatine surface—the thin cover glass was next the moving photographic plate and was only a small fraction of a millimetre distant from it. The breadth of the slit was about $\frac{1}{50}$ mm.

The carrier of the photographic plate was clamped to the middle portion of a wire stretched horizontally over two pulleys; a weight was attached to one end of the wire, while the other was attached to a piston, the motion of which in its cylinder caused oil to be driven through a fine hole in a brass disc. By turning the disc any one of a graduated series of holes could be brought into action according to the speed of travel desired.

The light from the source of illumination—a paraffin lamp—could be cut off momentarily by means of a shutter which was worked by a cord from outside the hut. In this way it was possible to record on the photographic plate the times of the beginning and ending of thunder. In the records reproduced (Plates 2 to 5) these momentary interruptions of the illumination are represented by vertical black lines; a single line indicates the beginning, a double line the end of a peal of thunder.

The interval on the photographic record between the vertical portion of the trace, which represents the sudden change of field due to the passage of a lightning flash, and the dark line which marks the moment when the thunder resulting from the flash began to be heard, afford data for obtaining an estimate of the approximate distance of the discharge.

The varying position upon the slit of the image of the meniscus on which the microscope is focussed is represented by the curve separating the dark and light regions of the record. The fine horizontal lines are due to dust particles or to irregularities of the slit; they are useful as reference lines from which the displacement of the meniscus may be measured. The vertical flutings which appear in some of the records are probably due to flickering of the lamp.

Records of the electrical effects of thunderstorms at various distances from the place of observation were obtained on ten different days in 1917. The records were not by any means continuous throughout the whole duration of a storm: comparatively quick runs of the recording apparatus were generally made—varying from 3 to 50 minutes in duration—and some time had to be spent in changing the photographic plates and readjusting the apparatus between the successive runs. Again, the difficulty of estimating the order of magnitude of the electric effects to be expected frequently led to the sphere being exposed when the test-plate would have been more suitable, or *vice versa*; the readings—which are 40 times larger with the sphere than with the test-plate—being in consequence too large or too small to be recorded. Thus the records obtained served rather to sample a thunderstorm at different stages of its history than to give a complete account of the changes in its electric field.

## II. *Some Typical Records.*

Enlargements of some of the records are reproduced in Plates 2 to 5. In the original negatives a change of one mm. in the ordinates represented a flow of 24 electrostatic units, or $8{\cdot}0\times10^{-9}$ coulomb through the electrometer: a change of potential gradient of 100 or 4000 volts per metre, according as the sphere or the test-plate was used, was required to cause the passage of this quantity of electricity between the exposed conductor and the earth.

A fairly typical fine weather record (May 23, 1917, 14h. 17m. to 14h. 51m. G.M.T.) is shown in fig. 1, Plate 2: the sphere was used as the exposed conductor. The record begins with a horizontal portion traced before the conductor was exposed to the electric field. The small peak near the beginning of the record shows the effect of raising the sphere to its maximum height (480 cm.), and immediately lowering it again into its protecting case; it indicates the existence of a positive potential gradient of 100 volts per metre. The sphere was raised at 14h. 20m., the exposure to the electric field being continued till 14h. 50m. except for regular interruptions at 5-minute intervals when the sphere was momentarily lowered into its case. The depths of the notches in the curve are measures of the potential gradient at the times of lowering the sphere: the potential gradients recorded at intervals of 5 minutes are, in volts per metre, 120 (at 14h. 30m. when the sphere was raised), 110, 120, 90, 75, 80, and finally 90 at 17h. 50m. (when the sphere was lowered).

The difference in the ordinates of the final and initial horizontal portions of the trace (both recorded while the sphere was in its case) is a measure of the integrated ionization current which has entered the sphere during the 30 minutes' exposure to the atmospheric electrical field. The record shows that this amounted to 21·8 E.S.U., while the mean charge induced on the earth-connected sphere during the exposure amounted to 23 E.S.U.—the equivalent of 96 volts per metre. The mean "dissipation factor" for the period of exposure was thus $\frac{21\cdot8}{30} \times \frac{100}{23}$, *i.e.*, about 3 per cent. per minute.

The readings obtained when the sphere is down form a series of points on a curve of which the vertical height above the initial horizontal part of the trace is a measure of the integrated ionization current which has entered the sphere from the atmosphere. This curve forms the zero line for potential gradient, *i.e.*, the differences of the ordinates of this curve and of the actual trace obtained when the sphere is exposed give a measure of the potential gradient at any moment.

In fig. 2 is reproduced the record of May 12, 1917, from 16h. 50m. to 17h. 35m. The sky was overcast and the weather conditions suggested thunder—a storm did in fact occur some hours later. The sphere was momentarily raised at 16h. 51m.; raised again at 16h. 55m., and kept up till 17h. 23m., being however momentarily lowered into its case at 5-minute intervals during this time; it was kept in its case after 17h. 23m. The potential gradient was 150 volts per metre at 16h. 51m.; it gradually diminished till it reached negative values, and continued to be negative from 17h. 12m. 50s. till 17h. 18m. 10s., reaching a minimum of −80 volts per metre at 17h. 16m., becoming positive again and being equal to 260 volts per metre when the sphere was lowered at 17h. 23m. The negative potential gradient coincided in time with the passage overhead of a cloud discharging rain which did not reach the ground. The test-plate was uncovered from 17h. 25m. to 17h. 30m.: the displacement of the meniscus on uncovering and covering the plate is almost too small to be seen in the reproduction of the record but indicates the continuance of a positive potential gradient of about 300 volts per metre. The ionization current from the earth-connected sphere to the atmosphere during the period of negative potential gradient has been sufficient to neutralise approximately the flow from the atmosphere to the negatively charged earth-connected sphere during its exposure to the positive potential gradient.

All the remaining records reproduced in the plates show the effects of lightning discharges (generally at a considerable distance) on the potential gradient.

Fig. 3 (June 13, 1917, 14h. 11m. to 14h. 16m. 30s.).

The sphere was exposed during the whole time represented by the record except at about 14h. 12m. 30s., when it was momentarily lowered; the effect of lowering and raising the sphere is indicated by the prominence midway between 14h. 12m. and 14h. 13m. The potential gradient at that moment was negative and equal to −420 volts per metre. The summit of the prominence gives the zero line of potential gradient. The record begins with a negative potential gradient of about −430 volts per metre. At 14h. 11m. 10s. distant electrical charges which were responsible for a portion (amounting to 150 volts per metre) of the negative potential gradient at the place of observation were neutralised by the passage of a lightning flash. The negative potential gradient at once began to be regenerated but was again suddenly diminished about 3 seconds later, losing 25 volts per metre by the passage of a lightning flash, probably at a still greater distance. This continuous production of a negative potential gradient and its sudden diminution at intervals by lightning discharges continues throughout the record. At about 14h. 13m. 40s. a sudden change of potential gradient of positive sign occurred, but was followed by one of negative sign and of nearly equal magnitude about 0·4 second later, a small positive change again occurring after another almost equal interval; these changes of potential gradient amount to +240, −220 and +25 volts per metre respectively. Another negative change of potential gradient (about 60 volts per metre) is indicated 10 seconds later. A few seconds after 14h. 16m. the record shows two discharges to have occurred with an interval of 2·4 seconds between them; each produced a change of potential gradient of positive sign, the first amounting to 840, the second to 870 volts per metre.

The potential gradient at any moment may be regarded as being the resultant of several electric fields, including those due to charges concentrated in different thunder-clouds or different centres of activity in the same cloud. The passage of a lightning flash results in the sudden destruction of one of these constituent fields. This at once begins to be regenerated by processes going on in the thunder-cloud at a rate which is indicated by the slope of the curve. The curve of recovery of the electric field (approximately logarithmic) shown after the discharges of 14h. 14m. is quite typical; similar curves appear in most of the records, a specially striking example being that of fig. 11 (Plate 4).

On account of the very short intervals between the successive peals of thunder, the times at which they began and ceased to be heard were not systematically recorded during the record reproduced in fig. 3. The first peal of thunder recorded is marked by the single and double dark lines as beginning at 14h. 13m. 8·9s. and ending at 14h. 13m. 15s., and a second one as beginning at 14h. 13m. 18·4s. and ending at 14h. 13m. 30s. The two peals are taken as being due to the discharges at 14h. 12m. 47·7s. and 14h. 12m. 53·6s. respectively; this gives a distance of 7·1 km. for the first and of 8·3 km. for the second. The first of these discharges produced a total change of potential gradient of 350 volts per metre, but this took place in two stages of 220 and 130 volts per metre which were separated by an interval of about 0·2 seconds; this interval is barely distinguishable in the reproduction. The discharge at 8·3 km. produced a change of about 95 volts per metre. The peal of thunder marked as beginning at 14h. 14m. 9s. probably belongs not to the flash at 14h. 14m., but to an earlier one, possibly the double one at 14h. 13m. 40s.

Fig. 4 (August 9, 1917, 14h. 45m. to 15h. 2m. 30s.).

Here the test-plate was exposed in place of the sphere. The potential gradient indicated at 14h. 45m. 15s. when the plate was first uncovered is negative (−4570 volts per metre). The principal sudden changes of potential gradient (all in the neighbourhood of 3000 volts per metre) are negative, indicating the destruction of positive fields by the passage of lightning discharges. The times of the beginning and ending of the peals of thunder were in most cases marked as shown by the single and double black lines. The distances indicated by the intervals between the principal discharges and the beginning of thunder are all about 5 km. The characteristic curve of recovery after the passage of each discharge is well shown.

Heavy black clouds were overhead at the beginning of the record and slight rain began about 14h. 48m. 30s. and became heavy at about the time when the record ceased. The effect of the rain is shown by the downward slope of the latter portion of the trace, which indicates a flow of positive electricity from the earth through the capillary electrometer to the test-plate. How much of this positive charge went to increase the induced positive charge on the test-plate (on account of increasing negative potential gradient) and how much to neutralise a negative charge carried down to the test-plate by rain drops, or by ions travelling under the influence of the negative potential gradient, remains undetermined owing to the fact that the cover was not replaced until after the record was completed.

Fig. 5 (June 17, 1917, 20h. 23m. 23s. to 20h. 27m. 29s.).

This is an enlargement of a portion of a record obtained while a severe storm was passing at a distance of 15 to 20 km. Between 20h. 20m. and 20h. 29m. the photographic trace recorded 95 positive discharges (*i.e.*, discharges causing a sudden positive change of potential gradient) and 40 negative discharges. The discharges were visible as vertical flashes passing between a cloud near the N.W. horizon and the earth, many of the flashes being multiple. The storm was seen to travel from W. to N.; newspaper reports show that it passed over St. Ives, which lies from 10 to 11 miles (about 17 km.) to the N.W., damage by lightning occurring there. The mean of the 95 sudden changes of potential gradient of positive sign amounted to 119 volts per metre, that of the 40 of negative sign to −80 volts per metre.

The sphere was used in obtaining this record.

Fig. 6 (June 16, 1917, 19h. 12m. to 19h. 23m.).

The test-plate was used as the exposed conductor.

The potential gradient was negative (−5400 volts per metre) at 19h. 12m. 45s. when the cover was removed, positive (=1000 volts per metre) at 19h. 22m. 15s. when the cover was replaced. Rain was falling throughout the duration of the record, and the charge carried down (by rain and ionization current) during the 9½ minutes' exposure was negative and amounted to $16 \times 10^{-12}$ coulombs per sq. cm., the mean current being thus about $27 \times 10^{-15}$ ampere per sq. cm. Two of the discharges recorded—at 19h. 17m. 4s. and at 19h. 20m. 55s.—were multiple, as is shown in the enlargements, figs. 18 and 19 of Plate 5. All the sudden changes of potential were negative—excepting the positive components of the multiple flashes—the largest amounting to −9600 volts per metre. The distance of this discharge, as is shown by the interval elapsing before the thunder began to be heard, was about 4·3 km. The distances of the others ranged between 4·3 and 5·7 km. The peals of thunder, as the intervals between the single and the double black lines show, were very long, some lasting for as many as 40 seconds.

Fig. 7 (June 16, 1917, 19h. 31m. 10s. to 19h. 36m. 45s.).

This is a portion of the record taken with the test-plate next after that shown in fig. 6. Rain continued to fall throughout the time of exposure. The cover was removed from the test-plate at 19h. 31m. 30s.; the potential gradient at that moment was negative, −6500 volts per metre; a lightning discharge had probably occurred immediately before the exposure of the test-plate. The discharge at 9h. 33m. 20s. was really multiple, the sudden changes of potential gradient being −2900, −5100 and +1300 volts per metre. The discharge at 19h. 35m. 55s. was negative (change of potential gradient = −4100 volts per metre) and was at a distance of about 5·5 km. The characteristic form of the recovery curve following the discharges is modified by the superposition of a general downward slope which represents an electric current from the ground to the atmosphere; this current was probably mainly carried by falling negatively charged raindrops. The potential gradient when the cover was replaced at 19h. 39m. (beyond the limits of the portion of the record reproduced) was positive and exceeded 2500 volts. The total quantity of electricity transferred per sq. cm. of the test-plate to the atmosphere, during the whole 8½ minutes of exposure, but mainly after 19h. 36m. was the equivalent of 40,000 volts per metre, *i.e.*, $3{\cdot}5 \times 10^{-11}$ coulomb.

Fig. 8 (May 29, 1917, 19h. 4m. 10s. to 19h. 11m. 50s.).

This is the final portion of a record which began at 18h. 44m. The sphere was used as the exposed conductor. The potential gradient had been +1200 volts per metre at 18h. 46m. when the sphere was raised; 980 at 18h. 51m. and −1400 at 18h. 56m., at which times the sphere was momentarily lowered. The peak shown in the figure at 19h. 6m. represents the effect of again momentarily lowering the sphere and indicates that the gradient was still negative, being equal to −430 volts per metre. At 19h. 11m. when the sphere was finally lowered the potential gradient had again become positive, being now +20 volts per metre.

These comparatively gradual changes of potential gradient accompanied the passage of towering cumulus clouds at no great distance. Superimposed upon them are sudden changes (amounting at most to 150 volts per metre) produced in the field by frequent discharges of more distant thunder-clouds. Some of these are positive, some negative; the discharges of either sign are alike in being followed by the characteristic curve of recovery of the field.

Fig. 9 (May 29, 1917, 17h. 58m. 45s. to 18h. 12m. 30s.).

This is an enlargement of a portion of a record which extended from 17h. 56m. to

18h. 36m.; the sphere was exposed. During the period covered by the portion of the record reproduced the sphere was momentarily lowered at 18h. and at 18h. 5m; it was also lowered at 18h. 10m. and kept in its case till 18h. 11m. when it was again raised. The potential gradient at the times of lowering the sphere amounted to +90, +60 and +40 volts per metre. It is plain from the record that the gradient remained positive throughout: the principal sudden changes of gradient were positive, and amounted to about 150 volts per metre; two, however, at about 18h. 1m. and 18h. 9m. 30s. were negative and equal to about 60 volts per metre. Positive discharges evidently also occurred during both the short periods for which the sphere was lowered.

The characteristic recovery of the field after both positive and negative discharges is well shown. The two peals of thunder recorded probably belong not to the discharges immediately preceding them but to the previous discharges. The discharges were probably at a distance of 20 km. or more.

Fig. 10 (August 15, 14h. 18m. 20s. to 14h. 30m. 15s.).

At 14h. 18m. 30s., when the cover was removed, the potential gradient was negative (=−3600 volts per metre). This negative potential gradient had increased to about −5000 at 14h. 19m. 6s.; at this moment the negative field was nearly destroyed by the passage of a lightning flash at a distance of 4·1 km., the sudden change in potential gradient being +4800. The five subsequent flashes also produced positive changes in the potential gradient; the beginning and ending of the thunder is marked on the record in each case. The magnitudes of the sudden changes of potential gradient vary from 3900 (the third shown in the fig.) to 14,600 volts per metre (the last); the distances of these two discharges were practically the same, 3·7 and 3·8 km.

The striking feature of this record is the abnormal character of the curve of recovery of the field after the passage of every discharge except the last; instead of the rate of recovery of the field being most rapid immediately after the discharge, it is at first zero or very small and gradually increases to a maximum, falling off again with the increasing field as in the normal type. The last discharge shown in fig. 10 as well as all the subsequent discharges of the record of which this is a part were followed by a recovery of the field of the normal type. The recovery curves following the discharges of the immediately preceding record of the same storm were also normal in character.

Rain began about 14h. 20m., became heavy about 14h. 25m., and ceased about 14h. 31m. 30s.

The potential gradient was negative throughout the period covered by fig. 10 until reversed by the last discharge shown. At 14h. 28m., when the cover was momentarily replaced, the potential gradient was − 4800 volts per metre. The small hump in the curve at 14h. 24m. 45s. is due to the shielding effect of a horse and cart which passed within a few yards of the test-plate.

In spite of the negative potential gradient, which would tend to produce an ionization current from the ground to the atmosphere, the total charge received by the test-plate from the atmosphere between 14h. 18m. 30s. and 14h. 28m. has been positive and equal to the charge which a potential gradient of 17,000 volts per metre would have induced on the earth-connected plate. The charge carried by the rain must thus have been positive and must have exceeded to the above extent the negative charge carried by the ionization current. The greater part of this charged rain has evidently fallen between 14h. 26m. and 14h. 28m.

Fig. 11 (June 12, 1917, 16h. 38m. 40s. to 16h. 50m.).

This is a portion of the second record taken on a sultry afternoon with towering cumulus in all directions. The first record ran from 15h. 55m. to 16h. 19m. A cap was seen to form on the summit of a large cumulus cloud in the E.N.E. at 15h. 59m., and another on one of the lower heads of the same cloud now in the N.E., about 16h. 16m. The potential gradient throughout this first record was positive and about 50 volts per metre. No thunder was heard and no sudden changes of the field are shown on the record.

The second record, of which fig. 11 is a portion enlarged, extended from 16h. 27m. to 17h. 4m. The large cumulus cloud was N. by E. with its edge at an elevation of about 60 degrees at 16h. 30m. and due N. about 16h. 50m. The potential gradient diminished from +44 volts per metre at 16h. 30m. to 15 at 16h. 35m. and then became negative, being −29 at 16h. 40m., −175 at 16h. 45m. (all the above being occasions of momentary lowering of the sphere), and −106 volts per metre at 16h. 49m. when the sphere was finally lowered. The field was zero at 17h. 14m., and had become positive at 17h. 16m., when the observations ceased.

All the sudden changes of field observed were positive; two of 18 and 14 volts per metre, both due to flashes at a distance of about 7 km., occurred before the field had become negative. The other two (equal to 120 and 320 volts per metre respectively) are shown in the figure; they both show characteristic recovery curves; in both cases the potential gradient was reversed, in the first by a discharge at a distance of 8·2, in the second by one at a distance of about 7 km.

The discharge at 16h. 45m. 50s. is an interesting one. The negative potential gradient had reached a steady value—about 170 volts per metre—before the passage of the discharge. The discharge—at a distance of about 7 km.—caused the gradient to become positive (=150 volts per metre); the negative field was again re-established, practically exponentially, a steady value being finally reached equal to about 105 volts per metre. The sphere was brought down at 16h. 49m. No thunder was heard after this time.

In Plate 5 are enlargements of small portions of some of the traces, showing details in the changes of potential gradient associated with lightning discharges. In each case the time in seconds is shown, reckoned from the moment at which the discharge, as indicated by the record, began.

The principal discharge of fig. 12 occurred on May 29, 1917, at about 15h. 23m. 10s.; the peal of thunder which followed began 21·5s. later (indicating a distance of about 7 km.) and was audible for about 20 seconds. The sudden change produced in the potential gradient was negative and exceeded 1250 volts per metre. The record shows the characteristic curve of recovery of the field, interrupted at 1m. 50s. after the discharge by the lowering of the sphere. The positive field due to the charge which the flash neutralised was nearly counterbalanced at the place of observation by a negative field, so that the resultant potential gradient before the passage of the discharge was only about +360 volts per metre.

When the sphere was first raised, at 15h. 19m. 30s., the potential gradient was positive—about 150 volts per metre—and it increased up to the moment of the principal discharge. There were, however, during this time, small sudden changes of potential, some positive, others negative, none exceeding 50 volts per metre; they were obviously due to very distant discharges; no thunder was recorded. Throughout the afternoon there were towering cumulus clouds in all directions, rain falling from some of them.

The discharge of fig. 13 occurred about 15h. 9m. 30s. on August 15, 1917. The sphere was lowered at 15h. 10m., the characteristic curve of recovery of the field being thereby interrupted. The peal of thunder began while the sphere was being lowered, *i.e.*, about 40 seconds after the discharge; the beginning is not marked on the record, but the double dark line indicates that the peal of thunder ended about 55 seconds after the discharge. The potential gradient immediately before the discharge had a negative value exceeding 1000 volts per metre; immediately after the discharge the potential was positive and equal to about 300 volts per metre.

The first discharge of fig. 14 occurred at 13h. 50m. on June 13, 1917, just at the moment when the sphere had been raised to its exposed position. The potential gradient before the discharge was negative (= − 690 volts per metre). The discharge was a double one, causing an increase in the negative potential gradient of more than 980 volts per metre, followed by a sudden change of the opposite sign, which brought the potential gradient to within 260 volts per metre of its original value, the total duration of the double discharge being about one-fifth of a second.

Two other double discharges of about the same total duration were recorded about 22 seconds and 87 seconds later, the first giving sudden potential changes of +70 and −30, the second of +100 and −115 volts per metre. The other discharges shown in the figure are noteworthy as not being followed by the usual recovery curve.

The next three figures are further examples of double discharge records of the same type—*i.e.*, of records showing the occurrence within a very short interval of time of two sudden changes of potential of opposite sign. They differ among themselves mainly in the relative magnitudes of the two sudden changes: the first change of gradient being the greater in fig. 15, the two being approximately equal in fig. 16, and the second being the greater in fig. 17; in the last case the initial change is negative, in the two others positive. The duration of the double discharge is about one-fifth of a second in fig. 15, two-fifths of a second in fig. 16, and two-fifths of a second in fig. 17.

Double discharges consisting of two successive sudden changes of potential gradient of the same sign are also not uncommon. A striking example is that of the last discharge shown in fig. 3, where a sudden positive change of potential of 840 volts per metre is followed 2·4 seconds later by a second change of the same sign amounting to 870 volts per metre. The discharge at 14h. 12m. 50s. (in the same fig. 3) was also really a double one of this type, the interval between the two components of magnitudes, 220 and 130 volts per metre, being about one-fifth of a second.

What have been called above double discharges, it should be noted, are not necessarily discharges along the same track or even from the same thunder-cloud; it may often be observed that lightning flashes from two different centres occur almost simultaneously.

In the last three figures of Plate 5 are reproduced enlargements of records of multiple discharges, *i.e.*, of records showing a rapid succession of changes of potential gradient of opposite sign. These were all obtained during the same thunderstorm, that of the afternoon of June 16, 1917. The first shows sudden changes of potential gradient of −9600, +4350 and −1500 volts per metre, the intervals between the reversals being about one-third of a second. The second shows sudden changes of potential gradiant amounting to −7100, +1700, −1700, +300, −1900, +700, −600, +1000, the total time occupied by the eight reversals being 2·1 seconds. In the third the changes of potential gradient are −1600, +900, −1600, +1200, +700, −700 volts per metre, the total duration being 1·9 seconds.

## III. *On the Prevailing Sign of the Sudden Changes Produced in the Potential Gradient by Lightning Flashes.*

The sudden changes produced in the potential gradient at the place of observation by the passage of lightning discharges have been more often positive than negative, *i.e.*, the greater number have consisted in a sudden increase of a previously existing positive potential gradient or a diminution or reversal of a previously existing negative gradient; in other words they might be explained as being due to the discharge of a negatively charged cloud. Discharges producing such a positive change of potential gradient are called in what follows positive discharges.

The number of positive discharges recorded in 1917 was 432, of negative discharges 279. If the observations of 1914 and 1915 are included, the numbers are 528 and 336, the ratio being 1·56. Of the ten days of thunder on which records were obtained in 1917, there were nine on which more positive than negative discharges were recorded; on the remaining day, however (June 16) about twice as many negative as positive discharges were recorded (74 negative, 38 positive). It is perhaps natural to associate the excess of positive over negative discharges with the excess of positive electricity found by SIMPSON* and others to be carried down in rain, the greater part of the charge transferred from the atmosphere to the earth by the rain of the thunderstorm being perhaps returned in lightning discharges. (See however Sections XIX. and XX.)

## IV. *Magnitude of the Changes Produced in the Electric Field by Lightning Discharges at Different Distances.*

The approximate distance of each lightning flash which caused a disturbance on the photographic trace was, when possible, determined by observing the time interval between the discharge and the thunder associated with it. The beginning of each

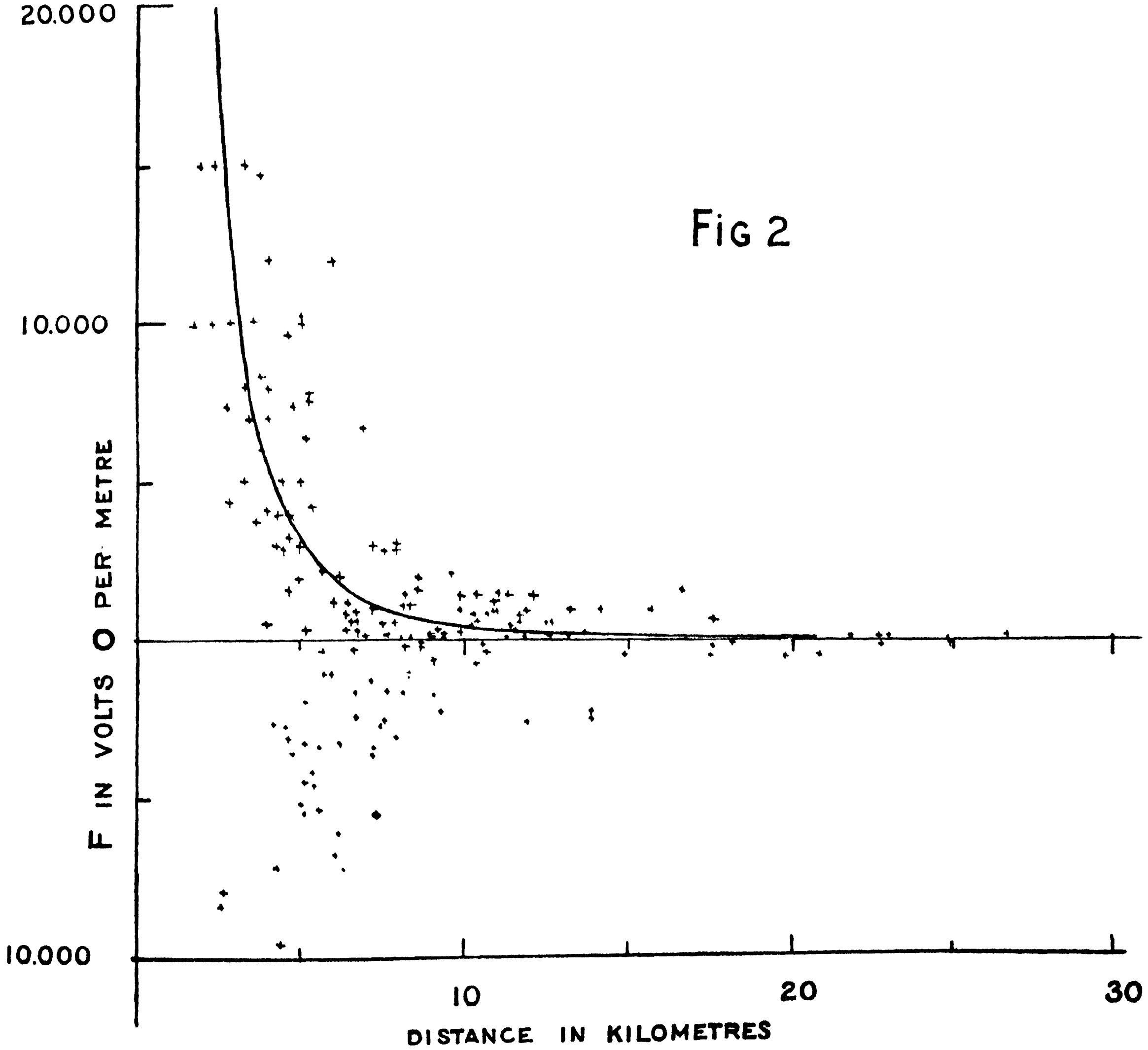


Fig 2

---

* SIMPSON, 'Phil. Trans.,' A, vol. 209, p. 379, 1909.

peal of thunder heard was marked on the trace by momentarily cutting off the light as described in Section I. It was by no means always possible to be certain which peal of thunder recorded was caused by the lightning discharge responsible for a given sudden disturbance of the field; when the storm was a distant one with very frequent lightning flashes there might be several subsequent discharges between the passage of a flash and the arrival of the sound of its thunder. There appeared to be no ambiguity in the case of about 120 discharges recorded in 1917; the approximate distance L of each of these discharges and the sign and magnitude of the resulting sudden change of field F are shown in fig. 2, which includes also the eye observations of 1914 and 1915. When the records show two or more sudden changes of field within a fraction of a second it is the largest of these which is recorded in fig. 2; it was considered that if the component discharges of the multiple flash were not all at the same distance, the one which produced the largest effect was likely to be the nearest, and therefore that of which the distance was deduced from the interval elapsing between the discharge and the beginning of the thunder.

## V. *Effects to be Expected from Different Kinds of Discharges at Different Distances.*

A lightning flash may consist in the passage of a charge Q from a certain region A of the atmosphere to earth, or from a region $A_1$ to another $A_2$ both in the atmosphere; $A_1$ and $A_2$ may be in the upper and lower parts of the same thundercloud with their centres near the same vertical line, or they may be at a considerable horizontal distance apart.

Let a charge Q derived from a certain region A of the atmosphere pass to earth. The change in the electric field may be considered as due to the removal of the charge Q from A and of an equal and opposite charge $-Q$ from A′ the image of A. Just as for many purposes no sensible error is made by assuming the magnetism of a bar magnet to be concentrated at two definite points, the poles of the magnet, so in the present case the charges Q and $-Q$ may be regarded as being concentrated at two points $p$ and $p'$. These points are situated at a height H above and at an equal depth below the surface of the ground, such that $2QH = 2\Sigma qh = M$, the electric moment of the discharge;[*] $q$ being the charge derived from a small element of volume at a height $h$. In calculating the change produced in the electric field at distant points by the passage of the discharge, no sensible error will be made by making this substitution, and even at points at no great distance from the axis $pp'$, the error will be small if there has been any approximation to a symmetrical distribution of the charge in a sphere surrounding $p$.

---

[*] In the present paper $2qh$, not $qh$ as in the previous paper, is taken as the electric moment of the discharge of a quantity $q$ from a height $h$ to earth, the moment with which we are concerned being that of charge $q$ at a height $h$ together with that of its image $-q$ at a height $-h$.

The vertical electric force at a point on the ground is given by

$$F = \frac{2QH}{L^3\left(1+\frac{H^2}{L^2}\right)^{\frac{3}{2}}} = \frac{2Q}{H^2\left(1+\frac{L^2}{H^2}\right)^{\frac{3}{2}}}$$

$\left(\text{becoming } \frac{2QH}{L^3} \text{ and } \frac{2Q}{H^2} \text{ when L is large and small respectively compared with H}\right)$ where L is the distance of the point from the axis *pp'*. The curve I in fig. 3 represents

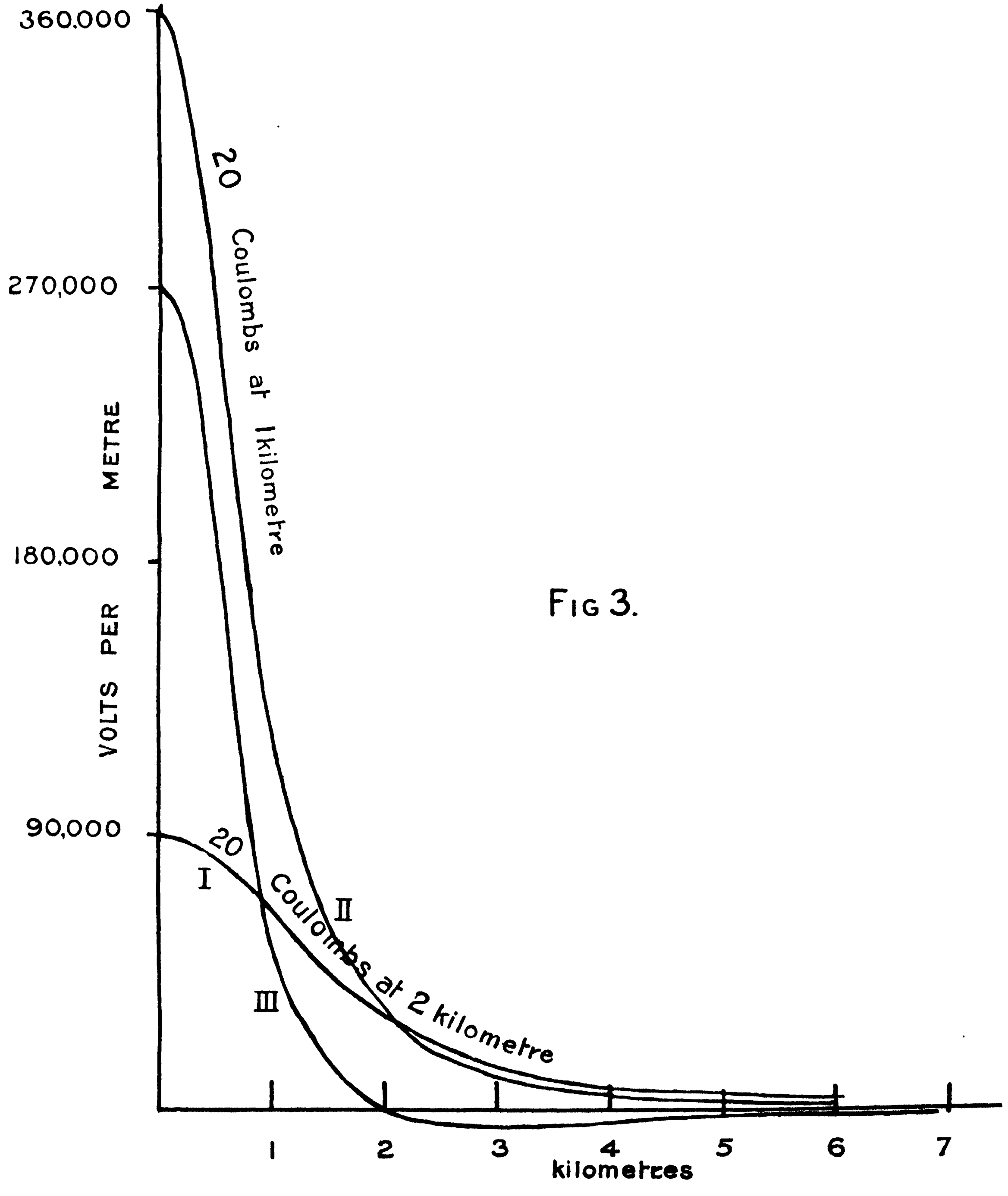


FIG 3.

the potential gradient at the earth's surface at different distances due to a charge Q supposed concentrated at a point at a height of 2 km.: it represents the change produced in the potential gradient by the discharge of 20 coulombs from a height of 2 km. to earth. The curve II. represents similarly the potential gradients due to the same charge at a height of only 1 km. The difference between the ordinates of the two curves (curve III.), represents the change of potential gradient produced by a vertical discharge of the quantity Q from a height of 2 km. to a height of 1 km. The sign of the effect is reversed at a certain distance, the vertical electric force at the surface of the ground being, for a given charge Q at the lower level, greater than for the same charge at the higher level when the distance is small, but less when the distance is great.

For a vertical discharge from a height $H_2$ to a height $H_1$, F, when L is large, becomes equal to $2Q(H_2-H_1)/L^3$, so that $FL^3 = 2Q(H_2-H_1)$ which may be defined as the electric moment of such a discharge. Thus when L is large $FL^3$ is equal to the electric moment of the discharge whether this reaches the earth or not.

The effects at different distances of various kinds of double discharge are also readily obtained from inspection of the two intersecting curves of fig. 3. For example, a discharge from a height of 2 km. to a height of 1 km., followed by an equal discharge from the lower level to earth, would produce at the surface of the ground two successive sudden changes of potential of the same or of opposite sign according as the distance exceeded or fell short of the above limit. Again if we consider a thunder-cloud of which the upper and lower parts are oppositely charged, and suppose that a discharge between the top of the cloud and the ground is followed by one between the ground and the bottom of the cloud, the two successive sudden changes of potential gradient would be of opposite sign, but their relative magnitude would depend on the distance of the place of observation from the discharges; at great distances the longer discharge, at small distances the shorter would produce the larger sudden change of potential gradient; while at some intermediate distance the two effects would be of equal magnitude. The various types of double discharge records shown in Plate 5 may perhaps be explained in this way; a given type of double discharge giving a considerable variety of effects on the trace according to its distance from the recording instrument.

If the effects of individual discharges could simultaneously be recorded at several suitable distributed stations, we should be able to learn much about the quantities of electricity which pass and about the initial and final distribution of charges. It is especially useful to have measurements of the change of field (1) at points at a considerable distance from a discharge, since the electric moment 2QH or $2Q(H_2-H_1)$ may at once be deduced, and also (2) for points nearly below the centres of the regions discharged, where, in the case of discharges to earth, F approximates to its maximum value $2Q/H^2$. Knowing both $2Q/H^2$ and 2QH we obtain both Q and H.

When a single station only is available we have to be content with attempting to learn something about the average lightning discharge by accumulating measurements of the effects produced by discharges at various distances.

## VI. *Electric Moments of the Discharges.*

For each discharge recorded in fig. 2, $FL^3$, the product of the vertical electric force and the cube of the distance of the discharge, has been calculated. This product, when the quantities are expressed in electrostatic C.G.S. units, may be taken as giving a lower limit for the electric moment $2\Sigma qh = 2QH$, or $2Q(H_2-H_1)$, becoming equal to it when L is large compared with H.

The mean value of $FL^3$ for the 78 positive flashes for which L could be determined in the 1917 records is $7\cdot3\times10^5$ in volts per metre × kilometres$^3$; for the 46 negative flashes the mean value obtained is identical with that found for the positive. We may take this value (the equivalent of $2\cdot4\times10^{16}$ E.S.U. × centimetres, or about 80 coulomb-km.) as a minimum estimate of the average electric moment of the lightning discharges.

The observations of 1917 give values of $FL^3$ ranging between 1/20 and 5 times the mean; but in more than half the discharges for which the necessary data are available $FL^3$ lies between one half and twice this mean value. Some of the eye observations made previously to 1917 lead to higher values, reaching in one case ten times the above mean.

In Table I. are given the mean values of $FL^3$ for positive and for negative discharges at distances (1) below 5 km., (2) between 5 and 10 km., and (3) exceeding 10 km. The number of observations used in getting the means is in each case inserted in brackets. $FL^3$ is given in volts per metre × kilometres$^3$ × $10^5$.

TABLE I.

| | Below 5 km. | | 5–10 km. | | Above 10 km. | |
|---|---|---|---|---|---|---|
| | Positive. | Negative. | Positive. | Negative. | Positive. | Negative. |
| 1917 only . . | 4·5 (17) | 3·7 (8) | 6·0 (32) | 8·2 (29) | 11·8 (29) | 9·5 (8) |
| 1914, 1915, 1917 . . . | 3·7 (37) | 3·7 (8) | 7·4 (48) | 7·8 (33) | 14·6 (38) | 15·5 (15) |

The mean values of $FL^3$ are not appreciably different for positive and negative discharges.

For discharges at distances between 10 and 15 km., the mean value of $FL^3$ for the 24 positive discharges of 1917 is $10\cdot8\times10^5$; if the 5 discharges of previous years are included the mean is 11·8 × $10^5$ volts per metre × kilometres³. Data for negative discharges between 10 and 15 km. are almost lacking.

These numbers leave little room for doubt as to the order of magnitude of the average electric moments of the discharges. Distances below 5 km. are too small in comparison with the probable lengths of the discharges for $FL^3$ to serve as a measure of the electric moment. We may assume that the value of $FL^3$ for a discharge at a distance of 10 km. or more approximates to its electric moment. The mean value of the electric moment for both positive and negative discharges may be taken as not differing much from $10^6$ in volts per metre × kilometres³ = 3 × $10^{16}$ E.S.U. × centimetres or about 100 coulomb-km.

Higher values for the mean electric moment are obtained, as is evident from Table I., if the data from discharges at greater distances than 15 km. are used. The records of discharges at great distances may possibly give disproportionately large values for the mean electric moment for two reasons: (1) because at these distances discharges of small electric moment are unrecorded on account of the small magnitude of the charges of potential gradient produced by them; and (2) because it is only at great distances that discharges, which do not reach the earth and which may be of great vertical length and have large electric moments, produce effects proportional to their moments. The sign of the effect of such discharges is in fact reversed at small distances, and the magnitude of the sudden change of potential gradient produced becomes more nearly proportional to the height of the lower end of the discharge than to its vertical length (fig. 3).

Some additions to the data of Table I. were furnished by the storm of June 17, 1917, in which the distance and frequency of the flashes were too great to admit of the distances of the individual discharges being estimated. There was in this case (see p. 80) independent evidence as to the approximate distance of the storm when the trace containing records of 95 positive and 40 negative discharges within 10 minutes was obtained. Assuming the distance of the discharges to have been 17 km. we obtain for the mean value of $FL^3$, in volts per metre × kilometres³, 5·8 × $10^6$ for the 95 positive discharges and 3·9 × $10^6$ for the 40 negative, corresponding to electric moments 2QH of 1·9 × $10^{16}$ E.S.U. × centimetres = 63 coulomb-km. and 1·3 × $10^{16}$ E.S.U. × centimetres = 43 coulomb-km. respectively. The discharges were observed to be approximately vertical and to pass between the base of the cloud and the earth.

## VII. *Quantity of Electricity Discharged in an Average Lightning Flash.*

When the electric moment of a discharge is known, the order of magnitude of the quantity of electricity which passes in the discharge may be roughly estimated. We may assume that the average vertical length of any ordinary discharge is likely to be between 1 and 5 km. Thus if the average electric moment 2QH is 100 coulomb-km., we may estimate the average quantity discharged in a flash as being between 10 and 50 coulombs.

We get some further information about the discharges by considering the way in which F varies with L (fig. 2). The charge which feeds a lightning flash is evidently not generally derived from a widely extended horizontal sheet, as is shown by the rapid falling off in F at comparatively short distances from the discharge.

The curve shown in fig. 2 represents the relation which would hold between F and L in the case of the discharge of 20 coulombs to earth from a point at a height of 2 km.; the charge may be considered to have been distributed symmetrically within a sphere around this point. The curve represents the mean of the observations fairly well, except in the case of discharges at great distances.

The average magnitude of the sudden change of field produced by lightning discharges at any distance may be roughly calculated by assuming that the average lightning flash consists of a discharge of 20 coulombs to earth from a height of 2 km.

The average change produced in the potential gradient by a discharge at a distance of 10 km. is, it will be noticed, of the order of 1000 volts per metre, and for moderate distances beyond this it probably falls off approximately according to the inverse cube law. (It should perhaps be pointed out that the change of field referred to here is merely the difference between the initial and final values, before and after the passage of a single discharge. At distant points the amplitude of the short period oscillations will greatly exceed the difference between the initial and final magnitudes of the field. Such oscillations—the ordinary "atmospherics" or "strays"—are of too short period to be recorded by the methods of this research).

Discharges may be expected to occur (1) between the ground and the lower part of a thunder-cloud; (2) between the upper and lower parts of the cloud; (3) between the upper part of the cloud and the ground; and (4) upwards from the top of the cloud. Great differences in the vertical lengths and in the electric moments of discharges are therefore to be expected, and the manner in which F varies with L in the different storms furnishes some evidence of such differences. When, as in the records of June 12, 1917, $FL^3$ varies little with the distance and is besides relatively small, one is tempted to conclude that the vertical length of the discharges was small, that, for example, they passed between the ground and the base of the cloud. When on the other hand, as on August 15, 1915, or August 15, 1917, $FL^3$ continues to increase with increasing distance and reaches very high values, great vertical lengths would appear to be indicated for the discharges. Possibly the discharges of greatest vertical length may be those between the top of a thunder-cloud and higher levels of the atmosphere.

It is unfortunate that no records were obtained of the effects of discharges from clouds immediately overhead; such observations of the maximum values of F would have given useful evidence bearing on the height from which the discharges came. A discharge of 20 coulombs from a height of 2 km. would cause at the ground a maximum change of potential gradient of nearly 100,000 volts per metre.

Comparatively few determinations appear to have been made of the dimensions of lightning flashes. A few are quoted by HANN,* the length of vertical flashes to earth generally ranging from 1 to 3 km. It is only rarely, in the photography of lightning, that the distance of the flash has been recorded, so that its length may be deduced. Fig. 4 is a reproduction of a photograph taken with this object in view and for which the necessary data are available; it is, moreover, of interest in other ways. It was taken on May 22, 1918, at about 22h. 45m., the camera pointing north. The interval between the lightning flash and the moment when the thunder began to be heard was 35 seconds, corresponding to a distance of 11·7 km. Two flashes are shown in the photograph, both passing between the cloud and the earth; they must have been nearly simultaneous, since the camera lens was covered as soon as a flash was observed. One discharge has initially passed upwards from the cloud and reached the ground by a curved path at a horizontal distance of nearly 4 km. from its starting point. The other has taken a nearly vertical course to the ground, its image is somewhat faint and ill-defined in the photograph: the discharge was probably within a heavy rain shower, a considerable thickness of which had to be traversed by the light on its way to the camera. The starting points of the two discharges in the cloud are comparatively close together, suggesting (as indeed does

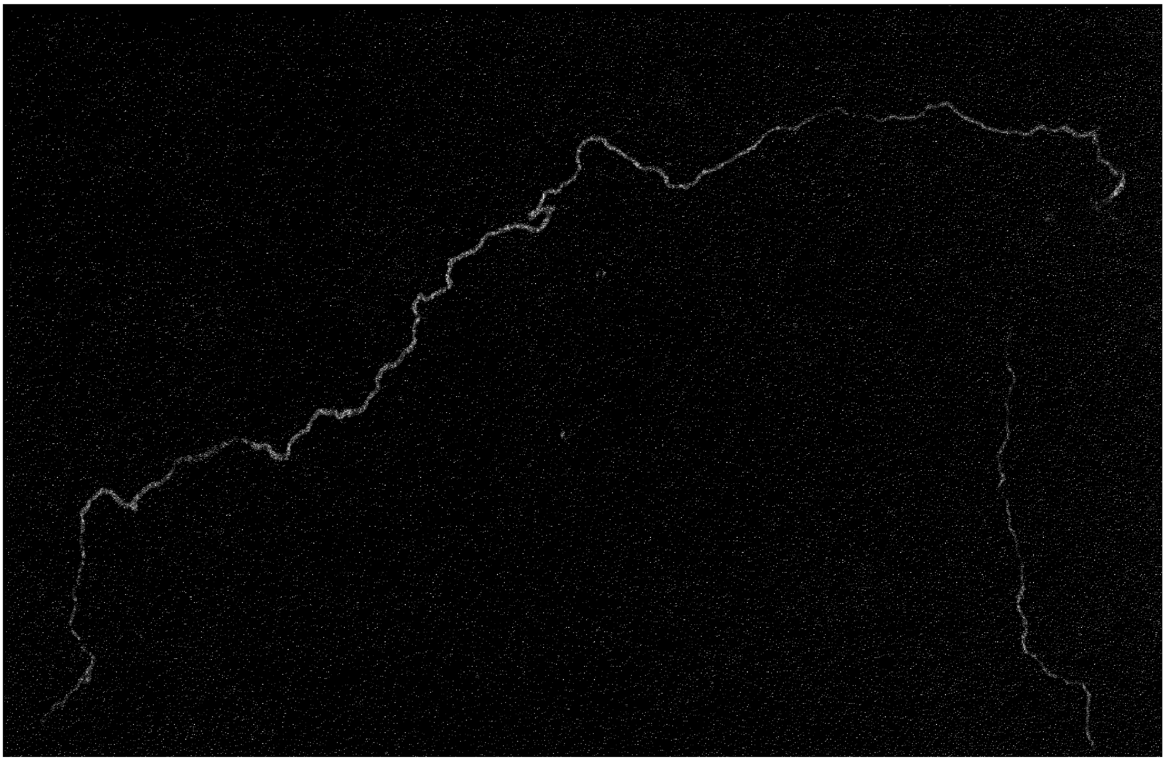


Fig. 4.

---

* HANN, 'Lehrbuch der Meteorologie,' p. 632, 1901.

the picture as a whole) that a charge of electricity had been concentrated in a comparatively small volume in the head of the cloud, and that the discharges took place approximately along lines of force.

The mean height of the upper ends of the two discharges—the height of the centre of the charged cloud-head according to this view—must have been just under 2 km., if its horizontal distance from the camera is taken as 11·7 km. The distance and height may in fact have been somewhat greater, since the track of the long flash may at some point of its course have been nearer the camera than the vertical flash, and the distance deduced from the interval between the lightning and thunder is that of the nearest point of the discharge.

## VIII. *Electric Field of a Thunder-cloud.*

It is much more difficult to obtain direct information about the electric field of a thunder-cloud than about the sudden changes produced in the field by lightning discharges. The observed field may be the resultant of the fields of several thunder-clouds superimposed upon the normal electric field; while a single instantaneous change in the field will in general be due to the passage of one lightning flash, of which the approximate distance may frequently be determined. Nothing approaching a direct survey of the electric field of a thunder-cloud has yet been attempted: some general conclusions may be reached by a study of the photographic records of the potential gradient in thunderstorms.

It might perhaps naturally have been thought that the actual field due to a distant thunder-cloud would greatly exceed in magnitude the sudden changes due to the lightning discharges from it, each flash removing from the cloud only a small part of its whole charge. This is disproved by the observations; only when there has been, in addition to the more distant thunder-cloud, a heavy shower-cloud overhead or in the immediate neighbourhood of the place of observation, has the actual potential gradient greatly exceeded the instantaneous changes; the main part of the observed field has in all such cases obviously been due to the nearer cloud and not to the comparatively distant thunder-cloud which was in action at the time. The potential gradient due to a distant thunder-cloud has apparently never greatly exceeded in magnitude the sudden changes produced in the field by the lightning discharges from the cloud. Very frequently each discharge has approximately destroyed or even reversed the previously existing potential gradient, the field has then been rapidly regenerated, to be again nearly neutralised or reversed by the next discharge. The magnitude of the vertical electric force at the ground due to a thunder-cloud at a given distance is thus probably of the same order as has been found for the change produced by the average lightning discharge at the same distance.

Potential gradients exceeding 30,000 volts per metre (*i.e.*, $\frac{1}{100}$ of the sparking value) have not been recorded: it is doubtful, however, if any of the records were obtained when the centre of a storm was nearly overhead.

There can be little doubt that it is by the agency of precipitation that the separation of positive and negative charges in a thunder-cloud and consequent production of an electric field is effected, the larger raindrops or hailstones carrying down a charge of one sign while the charge of opposite sign is attached to small drops or cloud particles carried up in the ascending air stream. It is not proposed to discuss here how the large and small particles may acquire charges of opposite sign: whether for example the thunder-cloud is essentially a frictional electrical machine (disruption of drops, which SIMPSON* regards as the important factor, being included under this head) or an influence machine as ELSTER and GEITEL† contend.

It is obvious that any view that places the seat of electro-motive force of a thunderstorm within the thunder-cloud implies that the cloud is essentially bipolar, equal and opposite charges being in any given time transferred from within the cloud to its upper and lower portions. The actual charges residing at any moment in the positive and negative portions of the cloud will in general be quite unequal, since the conditions determining the rates of dissipation of the charges at the top and bottom of the cloud will be very different; an important part of the loss of charge from the lower part of the cloud is obviously the charge carried down to the ground in raindrops. The lower charge may indeed to a large extent reside on rain-drops falling from the cloud, and may thus extend all the way to the ground. Rain may not however reach the ground, or may lose a large part or the whole of its charge before reaching it by processes to be considered later.

Consider a cloud in which there is an upward stream of charged cloud particles or small drops and a downward stream of oppositely charged large drops; the total vertical electric current is the sum of the currents carried by the upward and downward streams. If the density of electrification of the two streams were the same and uniform throughout the greater part of the vertical thickness of the cloud, then the whole of this portion of the cloud would be electrically neutral. Above a certain level however the small drops alone will remain, and again it is only the large drops which fall below the lower margin of the cloud; equal and opposite charges will in this way be liberated in the upper and lower portions of the cloud. The assumption of uniform density of electrification in the two streams is of course an extreme and improbable one, and the concentration of the charges in the upper and lower parts of the cloud alone is not likely to be so complete as this supposition would imply; it serves, however, to indicate the possibility of the positive and negative charges of a cloud being separated by a considerable vertical thickness of electrically neutral cloud.

The factors which determine the rates of dissipation of the upper and lower charges and the magnitudes of the maximum charges are considered in a later section. The electric field at the ground due to a cloud of this kind will be the resultant of the fields of the upper and lower charges.

---

* SIMPSON, *loc. cit.*

† ELSTER and GEITEL, 'Wied. Ann.,' 25, p. 116, 1885; 'Physikal. Zeitschr.,' 14, p. 1287, 1913; GEITEL, 'Physikal. Zeitsch.,' 17, p. 455, 1916.

In the ordinary thunder-cloud or cumulo-nimbus cloud we are concerned with rapidly ascending air currents of comparatively small horizontal dimensions. The heads of such clouds generally reach to heights of several kilometres: according to WEGENER* the top of a thunder-cloud may reach almost to the upper limit of the troposphere (about 10 km.). The average height of the bases is probably about 1 km.

If we suppose a cumulo-nimbus cloud to have charges $Q_2$ and $Q_1$ of opposite sign in its upper and lower portions, we may, for the purpose of calculating its electric field at a distance, treat these charges as if they were concentrated at definite "poles" at heights $H_2$ and $H_1$. The effect of the charges induced on the surface of the ground is the same as if they were replaced by charges equal and opposite to $Q_2$ and $Q_1$ and at depths $H_2$ and $H_1$ below the surface. The problem is then the same as that of finding the magnetic field due to two bar magnets of lengths $2H_1$ and $2H_2$ and moments $2Q_1H_1$, $2Q_2H_2$, placed so that their centres coincide, the axes being vertical and their polarities opposed.

The vertical electric force due to the cloud at a point on the ground at a distance L from the axis is

$$F = \frac{2Q_2}{H_2^2\left(1+\frac{L^2}{H_2^2}\right)^{\frac{3}{2}}} - \frac{2Q_1}{H_1^2\left(1+\frac{L^2}{H_1^2}\right)^{\frac{3}{2}}}.$$

Immediately below the cloud, where $L = 0$, the second term (representing the effect of the lower charge) will be the greater unless the ratio of $Q_2$ to $Q_1$ exceeds $H_2^2/H_1^2$, and for distant points the first term (representing the vertical force due to the upper charge) will be the greater unless $Q_1/Q_2$ exceeds $H_2/H_1$. Thus the surface of the ground may generally be divided into two areas, an inner and outer, in which the electric field due to the cloud has opposite signs; in the central area the effect of the lower pole of the cloud predominates and determines the sign of the potential gradient and of the charge on the ground, while in the outer area the effect of the upper pole of the cloud is the greater.

The maximum intensity of the resultant field anywhere near the centre of the inner area will generally greatly exceed the maximum reached in the outer area. The curve III., fig. 3, represents the resultant potential gradient produced at the ground by equal and opposite charges of 20 coulombs at heights of 1 and 2 km. The inner area has a radius of approximately 2 km., the maximum potential gradient at the centre amounts to 270,000 volts per metre, while the maximum reached by the potential gradient of opposite sign in the outer area is less than 10,000 volts per metre. Greater differences in heights of the two poles are probable in actual thunderstorms, and the difference in the intensities of the electric fields in the inner and outer areas is likely to be even greater than in the example given. As represented in fig. 5 lines of force from the central area end on the lower charged portion of the cloud, those from the outer area on the upper charge, others again connect the upper and lower charges.

---

* WEGENER, 'Thermodynamik der Atmosphäre,' p. 210.

Thus far no account has been taken of the conducting layer in the higher levels of the atmosphere, to the existence of which the phenomena of terrestrial magnetism seem to point.

The normal potential gradient at the surface of the ground in clear weather is of the order of 100 volts per metre, falling off with increasing height and becoming negligible above 10 km.; thus the potential in the conducting layer over regions of fine weather is not likely to exceed a value of the order of 1,000,000 volts. If we assume, in accordance with modern theories of terrestrial magnetism,* that the conductivity of the upper atmosphere is high enough to prevent any large potential differences within it, then even above a thunderstorm the potential in the conducting layer may not greatly exceed 1,000,000 volts. The potential in the head of a thunder-cloud probably reaches values 1000 times as great.

One important effect of the conductivity of the upper atmosphere is to cause a portion of the lines of force from the head of the thunder-cloud to end in the conducting layer. The effect will be more marked than that which would be produced by a solid conducting sheet since ions of opposite sign to the charge on the head of the cloud will be dragged down out of the conducting layer to form an expansion of it extending downwards towards the thunder-cloud. The charge on these ions (which constitutes the induced charge on this protuberance from the conducting layer) will partially neutralise the electric field produced at the ground by the charge in the head of the cloud; in other words lines of force from the head of the cloud which would otherwise have ended on the ground are now diverted upwards into the conducting layer.†

The considerations brought forward in this section suggest that the electric field of a cumulo-nimbus cloud may be regarded as due to charges, generally unequal, in the upper and lower parts of the cloud (falling rain being included as part of the cloud) and to the charges induced by these on the ground and on the conducting layer of the upper atmosphere. Thus the lines of force of the cloud may be divided into four classes, connecting (*a*) the ground and the lower charge of the cloud, (*b*) the lower charge and the upper charge, (*c*) the upper charge and the ground, (*d*) the upper charge of the cloud and the conducting layer of the atmosphere.

---

* SCHUSTER, 'Phil. Trans.,' A, vol. 208, p. 163, 1907; S. CHAPMAN, 'Phil. Trans.,' A, vol. 218, p. 1, 1919.

† In the absence of a previously existing conducting layer a thunderstorm would itself produce ionization in the upper atmosphere; this is readily seen to follow from the values found for the electric moments of lightning flashes.

If uniform stratiform conditions over a wide area be assumed, the conditions are simpler than in the case of the cumulo-nimbus cloud. The field at the ground below such a cloud, if the effects of the conducting layer be ignored, would be the difference between the fields due to the upper and lower charges, and its sign would be that of the field due to the larger charge. The effect of the conducting layer, as in the case of the cumulo-nimbus cloud, is to reduce the potential gradient produced at the ground by the upper charge of the cloud: firstly by the action of the opposing field of the charge induced on the conducting layer above the cloud, and secondly by the actual diminution of the cloud charge by the ionization current from the conducting layer.

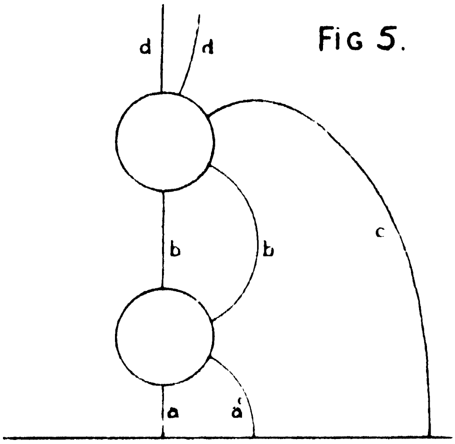


FIG 5.

## IX. *Conditions Determining Discharge.*

In order that a lightning discharge may begin, it is clear that the electric force must somewhere exceed the sparking limit, which amounts at the ordinary atmospheric pressure to about 100 electro-static units or 3,000,000 volts per metre; it is not necessary that the electric force along the whole length of the path of discharge should previously have approached the sparking limit. As LARMOR has pointed out,* if we suppose that an initial discharge occurs along a narrow line of length equal to the distance (possibly very small) over which the sparking value of the electric force was originally exceeded, and that this approximately equalises the potential along its path, there will be concentration of charge and intense local fields at the ends of this line; the discharge will thus be lengthened. The conditions are in fact momentarily much the same as if a conducting wire were placed along the path of this initial discharge. The maximum value of the electric force at the ends of the conducting track of the initial discharge will thus greatly exceed the critical

Sir JOSEPH LARMOR and J. S. B. LARMOR, 'Roy. Soc. Proc.,' A, vol. 90, p. 312, 1914.

value and will continue to do so as the track lengthens, so that the discharge may finally extend far beyond the boundary of the region in which the critical electric force was originally exceeded. Consider a stratiform cloud in which vertical separation of positive and negative electricity is taking place so that opposite charges are accumulating in the upper and lower portions of the cloud. Let us suppose that these charges remain approximately equal. There will be a vertical electric force within the cloud reaching a maximum in the central neutral zone of the cloud; the vertical electric force at the ground will be small and the conditions for discharge will be first reached within the cloud. Discharge will occur when the maximum vertical electric force within the cloud reaches its critical value; this amounts to about 30,000 volts per centimetre (= 100 electrostatic units) at a pressure of one atmosphere and is proportional to the pressure.

It is perhaps doubtful whether the vertical potential gradient within a cloud has necessarily to reach the above value of 3,000,000 volts per metre, in order that a discharge may begin, since an electric force amounting to one-third of this would be sufficient to bring the maximum electric force at the surface of a suspended drop, assumed spherical, to the above value. It must however be remembered that the critical value of the electric force at a curved surface of a conductor increases rapidly with the curvature and that only drops of the largest size will have any marked effect in assisting discharge.*

The discharge may extend considerably beyond the limits of the zone in which the vertical electric force originally reached the critical value. It is possible that it might extend even beyond the upper and lower boundaries of the cloud, for the ends of a linear vertical discharge would, as it lengthened, be continually penetrating into regions of a greater potential difference until they reached the limits of the charged portions of the cloud, so that the density of electrification and maximum electric force at the ends of the conducting track, and the consequent tendency to further lengthening of the discharge, would be increasing up to this point.

The end of an initial discharge which has penetrated into a region where there is little electric force to guide it will tend to branch or to expand into a brush. The potential may thus finally be approximately equalised throughout a considerable volume at each end of the discharge, the effective electric capacity of the expanded ends of the discharge and the original difference between the potentials of the regions thus connected will determine the quantity of electricity discharged by the complete flash.

If the lower charge of the stratiform cloud reaches nearly or quite to the ground

---

* The electric force at the surface of a conducting sphere 6 mm. in diameter has to reach about 260 E.S.U. (a value equal to nearly three times the sparking limit for a uniform field) in order that a spark may pass. (SCHUSTER, 'Phil. Mag.,' vol. 29, p. 182, 1890.) Drops of this size—which is little short of the maximum attained by rain drops—will only slightly assist discharge in an electric field in which they are suspended, and drops smaller than 3·5 mm. will not assist discharge at all.

(as will generally be the case when rain is falling), or if its charge is considerably smaller than the upper charge, then the initial discharge is likely to extend downwards till it reaches the ground; it will form then a conducting path for the main discharge, which may be regarded as reducing approximately to zero the whole discharge track and its ramifications.

If the potential gradient at the ground reaches a value amounting to any considerable fraction of that in the cloud, if for example, the upper and lower charges of a stratiform cloud are very unequal, then the critical value of the electric force is likely to be first reached, and the initial discharge to begin, at the surface of a projecting earth-connected conductor.

Let us next assume—and this perhaps represents more nearly the conditions which hold in an ordinary thunder-cloud—that the vertical separation of the centres of the charges is as great as the horizontal dimensions of the charged portions of the cloud.

Consider for example a charge to accumulate in the head of a cumulo-nimbus cloud until the conditions for the passage of a lightning discharge are reached. To get an idea of the order of magnitude of the quantities involved let us assume that the charge is distributed symmetrically about its centre within a sphere of radius R, the maximum electric force being at the boundary of the sphere. If the total charge of the sphere is Q, the radial electric force exerted at its surface by the charge is $Q/R^2$ and is there a maximum. A radial discharge will therefore begin at a point on the boundary of the sphere when F exceeds the critical value and will be continued inwards towards the centre and outwards approximately along a line of force. The charge of opposite sign in the lower part of the cloud will increase the electric force below and diminish it above the upper charge; the effect will however be small if the lower charge is small or if it is situated at a height small compared with that of the upper charge; in the latter case the effect of the lower charge is largely neutralised by the force due to its image. On the other hand on account of the diminished pressure at the greater height a smaller electric force is required to start a discharge from the upper than from the under side of the upper charged portion of the cloud.

Thus discharges may be expected to start not only downwards but also upwards and laterally from the charged head of a thunder-cloud. The path of discharge is likely to follow approximately a line of force which may belong to any of the classes of Section VIII. Discharges such as that of fig. 4, or even discharges upwards into a cloudless sky, such as have sometimes been observed, are not unlikely occurrences.

If an initial discharge from the charged head of a cloud reaches the ground, thus opening up a conducting path to earth, an approach to complete discharge is probable, so that the quantity of electricity which passes in the lightning flash may be taken as a measure of the charge which had accumulated in the head of the cloud.

A discharge originating in the region surrounding the lower pole of a cumulo-nimbus cloud is more likely to begin in the lower rather than the upper boundary of the charged region; since the electric force below will be increased, and that above diminished by the action of the induced charge, *i.e.*, virtually by the image of the lower charge. An extreme case is that in which the lower charge, carried largely by rain below the actual cloud, extends to the ground. Here the maximum value of the electric force would be at the surface of the ground; and, if the charge be assumed to be distributed uniformly throughout a region of which the vertical and horizontal dimensions are approximately the same, the maximum vertical electric force would not differ much from $2Q/R^2$, where R is the height of the centre of the lower charge. In this case the field will be locally intensified at the surfaces of projecting parts of earth-connected conductors, and discharges (not necessarily developing into lightning strokes) will occur from such points long before the electric force over flat ground reaches the sparking limit.

## X. *Dimensions of the Regions Discharged by Lightning Flashes.*

It has been shown that the quantity of electricity which passes in an average lightning discharge - if the thunderstorms investigated may be taken as typical—is of the order of 20 coulombs. In this and the following sections, X. to XVII., are considered some of the consequences which follow if the quantity discharged by a lightning flash is taken as 20 coulombs.

Consider a thunder-cloud of the bipolar type and assume that a discharge takes place when the electric force at the boundary of either the upper or the lower charge reaches the sparking limit $F_0$. To get an idea of the order of magnitude of the effects, let us assume that the charge is contained within a sphere of radius R, at a distance from the ground and from other charged masses, and that it is distributed symmetrically in such a way that the maximum radial electric force is at the boundary. A discharge will occur when $Q/R^2 = F_0$. Thus, if $Q = 20$ coulombs $= 6 \times 10^{10}$ E.S.U. and $F_0 = 30{,}000$ volts per centimetre $= 100$ E.S.U. (its value at ground level) then $R = 250$ metres. If $F_0 = 50$, its value at a pressure of half an atmosphere, $R = 350$ metres. If an equal and opposite charge (the other cloud-charge or the image of the first in the ground) were similarly distributed within a sphere of the same radius in contact with the first, we should have at the moment of discharge $2Q/R^2 = F_0$; and the values found for R would be $\sqrt{2}$ times as great as in the case considered, *i.e.*, 350 and 500 metres respectively.

A similar result is obtained if, instead of assuming the charge to have been distributed in a sphere, we suppose the vertical thickness of the charged portion of the cloud to have been small compared with its horizontal dimensions. Consider for example the case in which there are frequent flashes between the earth and the base of the cloud. We may picture the charged rain escaping from the base of the cloud as forming a charged layer which increases in thickness at a rate equal to the downward velocity of the drops. The vertical electric force at the upper and lower boundaries of the charged layer, due to its charge, will amount to $2\pi\rho d$ where $\rho$ is the charge per unit volume and $d$ is the vertical thickness. On this will be superimposed the electric force due to the upper charge of the cloud and that due to the induced charge on the ground; the first of these will increase the electric force at the upper surface and diminish that at the lower surface, while the second will increase the electric force at the lower and diminish that at the upper surface of the charged layer. If we assume the electric force below the lower charge to be greater than above it—as may easily be the case if the vertical thickness of the cloud (of cumulo-nimbus type) is great in comparison with the height of the lower charge—its magnitude will be between $2\pi\rho d$ and $4\pi\rho d$. A flash will occur when the vertical electric force reaches the sparking limit, *i.e.*, about 100 in electrostatic measure. If we assume the boundary of the lower charge to be a circle of radius $r$, and the quantity discharged to be 20 coulombs $= 6 \times 10^{10}$ E.S.U., $r$ is between 350 metres and 500 metres, these being the limits obtained by putting $F_0 = 2\pi\rho d$ and $F_0 = 4\pi\rho d$ respectively.

It has thus far been assumed that the horizontal dimensions of the charged portions of the cloud are less than the distance apart of their centres, and that the greater part of the whole upper or lower charge of the cloud is neutralised by each discharge. Let us now suppose that there has been a uniform stratiform distribution of charges over a wide area. Take as an example the cases in which the upper and lower charges of the cloud are equal, the other extreme case in which one of the charges is very small compared with the other is not essentially different—the charge on the ground taking the place of the second cloud-charge. There will be a discharge when $4\pi\sigma = F_0$, $\sigma$ being the total charge in a vertical column of unit area extending throughout the whole thickness of either charged portion of the cloud. If 20 coulombs are discharged in a lightning flash, and the whole thickness of a limited area of the charged portion of the cloud is discharged by the flash, the area A, discharged is such that $AF_0/4\pi = 20 \times 3 \times 10^9$; if the area discharged be assumed circular, and $F_0$ be taken as 100, the radius of the area discharged must be approximately 500 metres.

## XI. *Maximum Potential Attained before the Passage of a Lightning Flash.*

The potential at the surface of the sphere, considered in Section X., will immediately before discharge be approximately $Q/R = F_0R$; other terms being relatively small may be neglected in estimating the order of magnitude of the potential. The potential at the centre of the sphere will exceed that at the boundary, the excess lying between zero and $F_0R$—these being the values in the limiting cases (1) in which the radial electric force within the sphere is zero, the charge being confined to the boundary, and (2) in which the radial electric force within the sphere reaches everywhere the sparking limit. (The case of uniform distribution of the charge within the sphere is intermediate, the excess being $\frac{1}{2}Q/R$). The potential at the centre thus lies between $Q/R = F_0R = \sqrt{QF_0}$ and twice this value.

If Q = 20 coulombs = $6 \times 10^{10}$ E.S.U. and $F_0 = 50$ E.S.U. the potential at the surface of the sphere before discharge must reach $1{\cdot}7 \times 10^6$ E.S.U. = $5 \times 10^8$ volts.

We may take $10^9$ volts as giving the order of magnitude of the potential reached in a thunder-cloud before the passage of a discharge of 20 coulombs.

The order of magnitude of the potential required to cause a discharge remains the same even if the spherical distribution of the charge is departed from: the horizontal dimensions might, for example, considerably exceed the vertical so long as they did not much exceed the height of the charge above the ground.

Suppose next that there is a stratiform distribution of charges over a wide area, so that the lines of force are vertical. The conditions of discharge have already been discussed in Section IX.

If we assume that the mean vertical electric force along the whole length of the line of discharge initially approached the value $F_0$ (= about $3 \times 10^6$ volts per metre) and that this length is 2 km., the potential difference between the levels connected by the discharge must have been about $6 \times 10^9$ volts. But, as was pointed out in Section IX., the discharge may extend much beyond the regions in which the vertical electric force had originally attained the sparking limit $F_0$; the discharge might, for example, extend from the region of the upper charge of the cloud to the ground, although the electric field did not originally extend to the ground. The potential difference required to produce a vertical lightning flash 2 km. long from a cloud of this type may thus be considerably less than $6 \times 10^9$ volts, but it is not likely to be so small as $10^9$ volts.

## XII. *Mean Density of the Charge in a Thunder-cloud immediately before Discharge.*

If we assume that a charge of 20 coulombs is concentrated within a sphere 500 metres in radius, the charge per cubic metre is about 120 E.S.U.

In the case of a stratiform distribution of charges we have immediately before discharge $4\pi\sigma = F_0$ (Section X.). If uniform density $\rho$ be assumed for the charge throughout a layer of thickness $d$, then $\rho d = F_0/4\pi$ = about 8 E.S.U. If $d$ be taken as equal to 1 km., $\rho = 8 \times 10^{-5}$ E.S.U. per cubic centimetre (= 80 E.S.U. per cubic metre). Concentration of the charge within a smaller thickness is probable, with a corresponding increase in the density of the charge.

The mean density of the charges in thunder-clouds is thus likely to reach values of the order of 100 E.S.U. per cubic metre.

## XIII. *Charge Associated with* 1 *c.c. of Water.*

If the amount of water in the charged portion of a thunder-cloud were no greater than in ordinary clouds (about 4 gm. per cubic metre), the average charge per gramme of water would be about 25 E.S.U.; the force exerted on each gramme of water by the electric field where it approached the sparking limit, 100 in E.S. measure, would amount to 2500 dynes, *i.e.*, to more than twice its weight. As pointed out by SIMPSON,* 10 E.S.U. is the largest charge per cubic centimetre of water consistent with its falling in an opposing electric field of 100 E.S.U. (on one occasion rain actually was found by him to carry a charge exceeding 10 E.S.U. per cubic centimetre). In the same paper SIMPSON draws attention to the very considerable accumulation of water that must occur in thunder-clouds through lagging of the larger drops behind the uprushing air. Thus the charge per cubic centimetre of water does not necessarily reach the above high values: and indeed the electric force opposing the fall of the large drops associated with the lower pole of the cloud cannot, as a rule, exceed their weight, since it is by the fall of these drops that the field is maintained. But there will be less concentration of water on the smaller drops associated with the upper charge, and densities exceeding 10 E.S.U. per cubic centimetre in the upper part of the cloud are not unlikely.

The drops in the head of a thunder-cloud may thus in virtue of their mutual repulsion have radial velocities which near the boundary may be comparable with the terminal velocity which the drops would acquire if falling freely through the air. Drops of $10^{-3}$ cm. in radius would have a maximum radial velocity of a few centimetres per second: if the radius were as large as $5 \times 10^{-3}$, the charge per cubic centimetre of water remaining the same, the radial velocity would be of the order of 1 metre per second. The characteristic bulging form of the heads of a developing cumulo-nimbus cloud may possibly be partly due to mutual repulsion of the charged droplets.

## XIV. *Disruption of Drops by the Electric Field.*

It was shown by Lord RAYLEIGH† that a charged spherical drop must become unstable if $Q^2$ exceeds $16\pi a^3 T$, where $Q$ is the charge, $a$ the radius of the drop and $T$ the surface tension. If the charge per cubic centimetre of the water in the cloud is $\rho$ and is equally distributed among the drops, so that $Q = \frac{4}{3}\pi a^3 \rho$, then the spherical form will be stable so long as $\rho^2 a^3$ does not exceed $9T/\pi$, *i.e.*, about 225 in the case of water drops. The limit fixed in this way for the maximum charge per cubic centimetre of water, even for rain-drops as large as $\frac{1}{3}$ cm. in radius (for which it amounts to more than 70 E.S.U.), is too high to be of importance in the thunder-cloud problem.

Of much greater importance is the effect, upon the stability of the drops, of the electric field in which they are suspended, or in other words of the induced charges on the two halves of each drop.

If it is as a result of the electric force within or at the boundary of a cloud that a lightning flash occurs, then it becomes an interesting question whether under certain conditions disruption of the drops may not occur before the conditions for discharge are reached. ZELENY,* who has made a very interesting series of investigations on the stability of electrified liquid surfaces, found that in air at atmospheric pressure the potential required to cause a discharge from the surface of a drop of water at the end of a capillary tube exceeds, though only by a few per cent., that required to produce instability and disruption of the drops. He points out that it would follow from his experiments that a discharge of minute electrified drops, constituting an upward shower, would take place from the edges of the wet leaves of a tree in a thunderstorm, before the electric force at the surface of the tree reached the sparking limits.

---

* SIMPSON, *loc. cit.*

† RAYLEIGH, 'Phil. Mag.,' vol. 14, p. 184, 1882.

It seems not unlikely that under certain circumstances a similar process may occur in a cloud, droplets suffering disruption where the field approaches the sparking limit. Consider a developing cumulus head in which a charge is accumulating, and suppose that the radial electric force near the edge of the cloud-head reaches the value required to cause disruption of the drops before it reaches the sparking limit. The induced charges on the two halves of the drop will then be separated and will tend to travel in opposite directions along a line of force.

The magnitude of the induced charge on each half of a spherical drop of radius $a$ in a field F is $3\pi a^2 F/4\pi = \frac{3}{4}Fa^2$, and when F approaches the sparking value this will generally greatly exceed any resultant charge of the drop. The charge per cubic centimetre of water for each half of the drop $= \frac{3}{4}F_0 a^2/\frac{2}{3}\pi a^3 = 9F_0/8\pi a$; if $F_0 = 100$ and $a = 1$ mm. the charge per cubic centimetre of water for each half of the drop is 360 E.S.U. Thus if the original drop of 1 mm. in radius were divided into two oppositely charged halves, the force acting on each of the new drops would in a field of 100 E.S.U. amount to 36 times its weight.

If the drop is drawn out by the action of the field into an ellipsoidal or cylindrical form before disruption, the induced charges will be considerably greater. Separation of the charges by division of the drop will thus give rise to oppositely charged portions each having a charge much greater than that of the original drop. The two portions will tend to travel in opposite directions along the line of force with velocities greatly exceeding the original radial velocity of the drop from which they were derived.

The outward moving products of disruption of the drops in the head of a cumulo-nimbus cloud may possibly constitute "false cirrus." These particles are more likely to freeze than those constituting the original head of the cloud; (1) because the stretching of the water drop into a filament itself causes cooling; (2) the conversion of a water filament into an ice crystal is not accompanied by a large increase of surface, and one of the main obstacles in the way of the freezing of small drops is thus removed; and (3) the charged particles are still further cooled through being driven by the action of the field into the colder and drier air outside the original cloud.

Ice needles formed under the above conditions would not only be charged but also electrically polarised (the induced charges of the original conducting filament remaining when the filament freezes), and will thus tend to remain with their long axes parallel to the direction of the electric force. A study of the optical phenomena of "false cirrus" would be of interest in this connexion, as would also an experimental investigation of the effects of an electric field on super-cooled drops.

---

* ZELENY, 'Proc. Cambridge Phil. Soc.,' vol. 18, p. 71, 1914.

### XV. *Pressure Within a Charged Portion of a Cloud.*

The pressure within a charged cloud—like that within a charged soap bubble—must be less than the pressure outside. If the whole charge be supposed to lie near the surface of a sphere the analogy with the soap bubble is complete, and this case may be considered in finding the order of magnitude of the effect. The reduction of pressure within the cloud by the charge is $2\pi\sigma^2 = F^2/8\pi$ where $\sigma$ is the charge per unit area of the surface of the sphere and F is the radial electric force immediately outside. Just before the passage of a discharge $F = F_0$ = about 100 in electrostatic measure, so that $F^2/8\pi$ is about 400 dynes per square centimetre, *i.e.*, about $\frac{1}{2500}$ of an atmosphere.

If we consider the charge to be distributed uniformly in a horizontal layer of thickness which is small compared with its horizontal dimensions, the diminution of pressure midway between the top and bottom of the charged layer, due to mutual repulsion of the charged drops, is again $F_0^2/8\pi$ dynes per square centimetre.

### XVI. *Thunder Resulting from Sudden Contraction due to Loss of Charge.*

Thunder is generally regarded as entirely due to the sudden expansion of the air along the track of a lightning flash. It is evident however that the sudden contraction of a large volume of air (the contraction corresponding to an increase of pressure of some tenths of a millimetre of mercury) must furnish a by no means negligible contribution to the thunder which follows the discharge.

### XVII. *Energy Dissipated in Lightning Discharges.*

If we take the estimates arrived at above ($V = 10^9$ volts, $Q = 20$ coulombs) for the order of magnitude of the potential in the charged portions of a thunder-cloud immediately before the passage of a flash, and of the quantity discharged in the flash, we obtain for the order of magnitude of the energy dissipated in an average discharge, $\frac{1}{2}QV = 10^{10}$ joules $= 10^{17}$ ergs.

We may also arrive at an upper limit for the energy if we assume that the distribution of the charges is stratiform and that the vertical electric force is uniform and equal to $F_0$ throughout the height H through which the discharge extends. From the value found for the average electric moment, 2QH, since V must now be equal to $F_0H$, we have for the energy dissipated, $\frac{1}{2}QV = \frac{1}{2}QF_0H$, about $10^{11}$ joules.

The rate at which electrical energy would be going to waste in a storm in which one such discharge occurred in every 10 seconds would amount to $10^{16}$ ergs per second or 1,000,000 kilowatts. It is of interest to compare this with the total power which would be available if it were possible to catch the rainfall of a thunder-shower before it fell and utilise the water power thus stored. The rate of rainfall in a severe thunderstorm may reach values approaching 10 cm. per hour. The water power available if it were possible to catch the rain at a height of 1 km. would amount to $3 \times 10^{15}$ ergs per sq. kilometre per second. Thus a rainfall of the above amount over an area of about 3 sq. km. if intercepted at a height of 1 km. would furnish sufficient power to produce the required electrical energy. The total power available for the production of lightning flashes may obviously greatly exceed the above estimate based on the rainfall.

## XVIII. *Interpretation of "Recovery" Curves.*

In a typical record of the changes of the vertical electric force due to a distant thunderstorm each vertical portion of the trace—representing the sudden change produced by a discharge—is followed by a characteristic "recovery" curve. This may be interpreted as follows:—The charge in the head or base of the thunder-cloud—or in both—is suddenly destroyed by the passage of a lightning flash. The field at once begins to be re-established at a rate represented by the initial steepness of the curve immediately after the discharge. But as the charge increases, its field tends to diminish the rate of increase of the charge in two ways: (1) by hindering the separation of oppositely charged rain-drops and cloud particles; and (2) by producing an ionization current which tends to neutralise the charge and increases with the increasing intensity of the field. Unless the field previously reaches the sparking limit, a steady condition will finally be approached when the two opposing processes, which tend respectively to increase and diminish the field, balance one another.

The initial rate of increase of the field immediately after the passage of a distant discharge is thus an important quantity. It is proportional to the rate at which a charge destroyed by the flash is regenerated by the action of the thunder-cloud, *i.e.*, it is proportional to the vertical electric current which is carried through the thunder-cloud by the convection of charged masses. If the distances and height of the charge destroyed are known, the vertical electric current may at once be deduced from the initial rate of increase of the field. If this information is not available the ratio of the current to the quantity which passed in the previous discharge can always be obtained from the record.

The value of $T = F \Big/ \dfrac{dF}{dt}$, where F is the instantaneous change recorded and $dF/dt$ the initial rate of recovery immediately after the discharge, has been deduced from the recovery curves in the case of 34 discharges. We may regard T as the time which would have been required to re-charge the cloud to the sparking limit had there been no neutralising process due to the action of the electric field of the cloud. The values of T vary between 1·5 seconds and 30 seconds, the mean of 64 measurements giving 6·9 seconds; in more than half the cases examined T lies between 4 and 10 seconds. These times are generally only a small fraction of the actual intervals between the flashes: on June 17, however, in a record (Plate 3, fig. 5) showing more than 100 flashes in 10 minutes—so that the average interval between the flashes was less than 6 seconds—the average value of T exceeded half this interval.

Some of the recovery curves, as, for example, that of June 12, shown in Plate 4, fig. 11, approximate very closely to the exponential form, so that the charge which has been regenerated when a time $t$ has elapsed after the discharge may be represented by $Q = Q_0(1 - e^{-\lambda t})$. Such a curve suggests that the charge of the thunder-cloud is being regenerated at a constant rate, and that it is at the same time being dissipated at a rate which is at any moment proportional to the charge. It might also however be interpreted as representing the regeneration of the charge by a constant E.M.F. in the cloud, the current through the cloud being proportional to the difference between this E.M.F. and the opposing potential difference produced by the charges separated; there would be no current when the charges reached a steady value. If dissipation of the accumulated charges is taken into account the recovery curve still remains of the same type; if the dissipation is large, or, in other words, if the internal resistance of the thunder-cloud, regarded as a generator of constant E.M.F., is large compared with that of the external circuit, the current through the cloud is constant, and we have again the case first considered.

The rate of regeneration of charge per second, in other words the vertical current through the cloud, immediately after a discharge varies between $\frac{2}{3}$ and $\frac{1}{30}$ of the charge removed by the flash, the mean being about $\frac{1}{7}$. If we assume a discharge to convey a quantity of the order of 20 coulombs, the mean current through the cloud, immediately after a discharge, is of the order of 3 ampères.

It is not at all impossible that this is also the order of magnitude of the vertical current through a thunder-cloud at other times than immediately after a lightning discharge, and even when an approximately steady condition of the field has been reached. Consider, for example, the charge in the head of a thunder-cloud which reaches to a great height. The conductivity of the atmosphere has been found by GERDIEN and by WIEGAND* to increase rapidly with the height, the former having found at 6 km. a conductivity more than 20 times as great, and the latter at 8865 metres a conductivity about 40 times as great as the normal conductivity near the ground. A charged body suspended in the atmosphere under the conditions found by **Wiegand** at 8865 metres would lose about $\frac{1}{15}$ of its charge per second. Thus a charge of 20 coulombs in the head of a thunder-cloud at this height should lose more than 1 coulomb per second: to keep the charge constant the vertical current through the cloud would have to exceed 1 ampère. The presence of such a large charge would, it is true, not leave the conductivity of the surrounding atmosphere unaltered: it would tend to increase it by dragging down ions from upper layers of still greater conductivity.

---

* WIEGAND, 'Deutsch. Physik. Gesellschaft,' February 29, 1914.

## XIX. *Electrical Currents Maintained in the Atmosphere by Thunder-clouds and Shower-clouds.*

Consider a cumulo-nimbus cloud of the type imagined in Section VIII. containing upper and lower charges—the latter being partly or, it may be, mainly carried by the rain below the cloud. Such a cloud may be regarded as an electric generator—whether essentially of the frictional type or of the influence machine type need not at present be discussed—capable of maintaining a potential difference between its poles of the order of $10^9$ volts.

As pointed out in Section VIII. the potentials in the conducting layer of the upper atmosphere is likely to be insignificant in comparison with that in the head of a thunder-cloud, and the potential difference between them may thus be of the order of $10^9$ volts.

There will be a flow of electricity along the lines of force belonging to the various groups enumerated in Section VIII. and indicated in fig. 5. The upper pole will continually be losing charge by currents flowing (1) to the lower pole; (2) to the earth's surface (this portion of the current reaching the outer zone (Section VIII.) where the potential gradient is unlikely to reach high values); and (3) to the upper atmosphere.

Unless the field in the shower-cloud approaches very near to the sparking limit, the conductivity within the cloud is likely to be small, since any ions liberated soon lose their mobility by becoming attached to cloud particles. The electrical resistance of the atmosphere between the upper pole of the cloud and the conducting layer of the upper atmosphere will be much less than that between the upper pole and the earth's surface; for the free ions will be dragged out of the conducting layer, and their mobility throughout the greater part of their course will greatly exceed that of the ions in the lower layers of the atmosphere. A large part of the current from the upper pole must thus go to the upper atmosphere.

Consider now the lower oppositely charged pole of the cloud. Part of the charge is continually being neutralised by the direct return current between the poles, but this, as has already been pointed out, is likely to be small. The greater part of the charge lost by the lower pole will reach the ground. If no rain reaches the ground the loss of charge will be due to ions moving under the action of the electric field of the cloud. If the normal rate of production of ions in the air below the cloud had alone to be taken into account, the current would be small; but we have to add the ions supplied by evaporation of charged drops falling from the cloud and those (of opposite sign) due to point discharges from earth-connected conductors, such as the leaves of trees or even the tips of blades of grass, under the action of the intense electric field of the central area below the cloud. If the rain reaches the ground the former of these sources of ionization is absent, but there is a further source of ionization in the splashing of the rain on the ground. In addition to the ionization current we have also the convection current carried to the ground by charged rain-drops. The total current between the lower pole of the cloud and the ground now consists of the convection current carried by the falling charged drops and the conduction current carried mainly by the upward stream of ions set free by point discharges and splashing at the surface of the ground. The ratio of the convection current to the conduction current will be less near the ground than higher up, since the falling drops will lose more and more of their charge as they penetrate farther into the stream of upward moving oppositely charged ions; these again as they are carried upwards by the electric field are continually diminished in number by union with the drops. The greater the supply of ions from the ground the smaller will be the charge retained by the drops; if the current carried by the upward stream of ions is sufficient, the drops may lose the whole of their charge or even have it reversed before they reach the ground. The charge carried to the ground by rain-drops is thus by no means necessarily a true measure of the vertical current in a shower: nor does the sign of the charge carried by the drops when they reach the ground necessarily indicate the sign of the current between the ground and the base of the cloud.

Thus a large part of the current from the upper pole of a cumulo-nimbus cloud is likely to reach the conducting layers of the upper atmosphere, while that from the oppositely charged lower pole goes mainly to earth. A current is thus maintained from the earth through the cloud to the upper atmosphere or in the reverse direction according to the sign of the polarity of the cloud.

Discharges between the ground and the lower pole of the cloud and between the upper pole and higher portions of the atmosphere contribute to the total current between the ground and the upper atmosphere; discharges between the two poles or between the upper pole and the ground diminish the electric field which maintains the vertical current without contributing anything to the current.

## XX. *Differences Between the Electrical Effects of Shower-clouds of Positive and Negative Polarity.*

We may define the polarity of a shower-cloud as being positive when the upper charge is positive, negative when the upper charge is negative, the current through it being upward in the former case, downward in the latter.

It was first proved by SIMPSON,* and has been confirmed by many observers, that rain on reaching the earth's surface is much more often positively than negatively charged. This, as we have seen, does not necessarily imply that shower-clouds are always or even prevailingly of negative polarity. It is therefore of interest to consider some of the differences to be expected between the electrical effects of clouds of positive and of negative polarity.

Recent experiments have shown† that the carrier of negative electricity in hydrogen, helium and nitrogen even at atmospheric pressure is the free electron, and that its mobility is some hundreds of times that of the carrier of positive electricity, the positive ion. In ordinary atmospheric air, as the pressure is reduced, the average mobility of the carriers of negative electricity increases relatively to that of the positive ions; quite an appreciable proportion of the negative carriers, consisting, according to WELLISCH,‡ of free electrons when the pressure is reduced to 8 cm. of mercury, the proportion increasing rapidly as the pressure is further reduced.

Thus, while the carriers of positive electricity dragged out of the conducting upper atmosphere by a cloud of negative polarity consist of ordinary ions, the negative carriers dragged down by a cloud of positive polarity are originally to a large extent free electrons, and a considerable proportion are likely to remain in this condition till quite moderate elevations are reached. The conductivity of the air between a shower-cloud and the upper atmosphere will thus be considerably greater if the cloud is of positive than if it is of negative polarity.

Let us compare two shower-clouds which differ only in the sign of their polarity and consider the effect of the greater conductivity of the atmosphere above the cloud of positive polarity. Let us suppose that the two clouds act as generators capable of maintaining equal potential differences between their poles. Let $V_2$, $-V_1$ be the potentials of the upper and lower poles of the cloud of positive polarity, and $-V_2^1$, $+V_1^1$ the potentials of the upper and lower poles of the cloud of negative polarity, let $V_2 - V_1 = V_1^1 - V_2^1$. Then the current from the ground to the upper atmosphere maintained by the cloud of positive polarity will be greater than that from the upper atmosphere to the ground maintained by the cloud of negative polarity, since the total resistance of the circuit is less in the former case.

The ratio $V_2/V_1$ is less than $V_2^1/V_1^1$, the upper and lower potentials being proportional to the resistance of the portions of the circuit above the upper and below the lower pole respectively.§ Thus $V_2^1$ is greater than $V_2$ and $V_1$ is greater than $V_1^1$; in other words the potential (and charge) of both the upper and the lower pole is greater when negative than when positive.

---

* SIMPSON, *loc. cit.*

† FRANCK and POHL, 'Verhandl. Deutsch. Physik. Gesellschaft,' 9, p. 69, 1907.

‡ WELLISCH, 'Phil. Mag.,' vol. 34, p. 33, 1917.

§ The potential of the conducting layers of the upper atmosphere is assumed to remain small in comparison with the E.M.F. of the thunder-cloud.

The potential gradient (negative) in the central area below the cloud of positive polarity will be greater than the positive potential gradient in the corresponding area below the cloud of negative polarity, the central positively charged area below the cloud of positive polarity being also larger than the negatively charged area below the cloud of negative polarity. Again, the positive potential gradient at the ground in the outer zone will be less (on account of the smaller charge on the upper pole) in the case of the cloud of positive polarity than the negative potential gradient in the corresponding region due to the cloud of negative polarity. Thus in each area negative potential gradients tend to be greater than positive.

The electric field in the central area below the lower pole being stronger in the case of the cloud of positive polarity, the current carried by the stream of positive ions from the ground will be increased, and therefore also the tendency to neutralisation or reversal of the negative charge on the falling rain-drops.

If lightning discharges occur, they are more likely to pass between the ground and either the upper or the lower pole if this is negative than if it is positive, since the charge of the pole is greater when negative. Thus discharges carrying positive electricity from the earth to the atmosphère will be more frequent than negative discharges. Discharges will tend to occur especially between the ground and the upper, negative, poles of clouds of negative polarity and the lower, negative, poles of clouds of positive polarity. In the latter case the discharges are an additional source of loss or reversal of the negative charge on falling rain-drops.

Essentially similar results are reached if, instead of assuming the same potential difference to be maintained between the poles, whether the clouds are of positive or of negative polarity, we assume that the same vertical current is maintained in both cases.

Thus, if we assume that shower-clouds may have polarity of either sign, the differences in the mobilities of the positive and negative carriers of electricity in the higher portions of the atmosphere will account for the preponderance in showers: (1) of negative potential gradients; (2) of upward or positive lightning discharges; and (3) of positively charged rain. It also affords (4) a possible explanation of the normal positive potential gradient of fine-weather regions.

## XXI. *The Normal Potential Gradient and Air-earth Current of Fine Weather.*

A thunder-cloud or shower-cloud is the seat of an electromotive force which must cause a current to flow through the cloud between the earth's surface and the upper atmosphere. In the case of thunder-clouds the records of the changes produced in the electric field by the passage of lightning flashes give us means of forming some estimate of the magnitude of such currents, and it would appear from them that the current through a few square kilometres of the surface of the ground below a thunder-cloud may amount to some ampères. In shower-clouds in which the potentials fall short of what are required to produce lightning discharges, there is no reason to suppose that the vertical currents are of an altogether different order of magnitude. If any considerable proportion of shower-clouds are of positive polarity the upper atmosphere will receive an excess of positive electricity which may possibly be sufficient to maintain the positive potential of the conducting layers and to supply the normal downward current of the fine-weather regions. The total current which must be supplied for this purpose is, as Simpson[*] has pointed out, of the order of 1000 ampères for the whole earth.

It is not necessary to suppose that only isolated clouds of the cumulo-nimbus type contribute to the current between the ground and the upper atmosphere. If we consider a cloud from which heavy rain is falling and assume that the conditions are uniform over a large area, the case is in fact somewhat simpler than that of the cumulo-nimbus cloud; the general results are the same.

We may suppose that a steady condition is reached in which the vertical electric field within the cloud (and thus the potential difference between its upper and lower surfaces) has a value which depends on the rate of rainfall and other factors; it is assumed to be independent of the sign of the polarity. Even if this potential difference is only $\frac{1}{10}$ or $\frac{1}{100}$ of that reached in thunder-clouds the effects may be important: the E.M.F. which tends to drive a current between the ground and the upper atmosphere is still from 10 to 100 times the normal potential of the upper atmosphere above fine-weather regions.

The difference between the mobilities of the positive and negative carriers dragged out of the conducting upper atmosphere will again cause clouds of positive polarity to differ from those of negative polarity in (1) the greater magnitude of the vertical current (positive for the cloud of positive polarity); (2) the smaller magnitude of the potential (positive) at the upper surface and greater magnitude of the potential (negative) at the lower surface of the cloud; and thus (3) the greater intensity of the potential gradient (negative for the cloud of positive polarity) below the cloud, this again tending to cause a larger part of the vertical current below the cloud to be carried by ions liberated at the ground and thus to produce a more complete discharge of the (negatively-charged) rain.

## XXII. *Influence of the Nature of the Earth's Surface below a Thunder-cloud or Shower-cloud.*

The dissipation of the lower charge of a thunder-cloud or other rain-cloud by the upward stream of ions liberated by point discharges or by splashing at the earth's surface must depend largely on the nature of that surface, on whether for example it consists of desert, snowfield, grassland, forest, lake or sea; and again the effect of the nature of the covering of the earth's surface may depend on the sign of the electric field.

---

[*] Simpson, 'Nature,' December 12, 1912.

Point discharges will occur most frequently and give rise to the largest currents over forests and lands covered with vegetation ; also on mountain summits and ridges, owing to the increased intensity of the electric field through proximity to the charged cloud. Ionization by splashing of rain on the ground and the relative number of positive and negative ions liberated thereby is likely to be very different over the various surfaces. Of special interest is the question of the amount and nature of the ionization at the surface of the ocean under heavy rainfall.

Over an area in which the surface ionization was large we should expect an increased vertical current, a diminution of the charge carried to the ground by rain, a diminution in the intensity of the electric field of the cloud, and in consequence a diminution in the tendency for lightning discharges to occur.

The holding up of charged rain-drops by the electric field and the diminution of the field by the dissipation of cloud charges by forests and other sources of surface ionization are perhaps not negligible factors in the local distribution of rainfall. Mr. L. F. RICHARDSON,[*] describing some of the phenomena observed during the passage of a line squall in France, on September 6, 1917, remarks "the cloud was noticeably darker over the Forest of Argonne than over the grasslands of Champagne."

## XXIII. *Secondary Thunder-clouds.*

It has thus far been assumed that the source of E.M.F. is within the cloud in which the lightning discharges and other electrical effects are manifested. It is easy however to imagine conditions in which a cumulo-nimbus cloud, which acts as electric generator, may supply electrical energy to quiescent clouds and produce in them intense electrical fields and even lightning discharges.

Consider for example a horizontal stratiform cloud which intersects lines of force connecting the poles of a cumulo-nimbus cloud; the stratiform cloud might be a lateral extension of the shower-cloud. The electrical conductivity within the stratiform cloud will, through capture of the ions by cloud particles, be very small compared with that of the free air above or below the cloud. The current from the poles of the primary thunder-cloud will cause an accumulation of charges of opposite sign at the upper and lower surfaces of the stratiform cloud. This accumulation of charge will continue—unless the field within the cloud previously reaches the sparking limit—until a steady condition is reached, when the vertical field within the cloud is sufficient to maintain a current equal to that which enters its upper and lower surfaces. The potential difference finally existing between the upper and lower surfaces of the stratiform cloud might amount to a considerable fraction of that between the top and bottom of the shower-cloud, the ratio being that of the resistance of that portion of a tube of flow which lies within the cloud to the resistance

---

[*] RICHARDSON, 'Quart. Journ. Roy. Meteor. Soc.,' 45, p. 112, 1919.

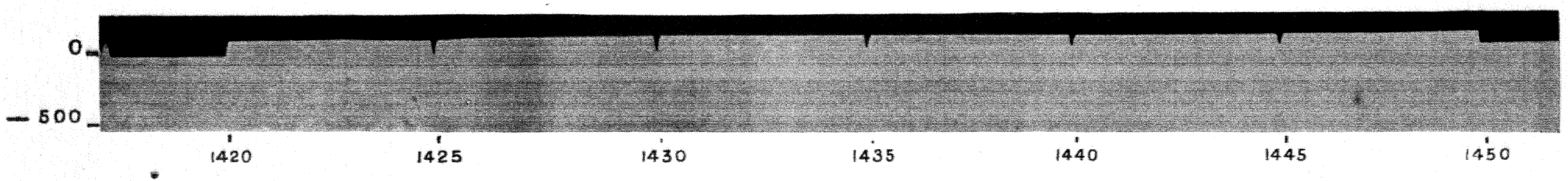


Fig. 1, May 23, 1917.

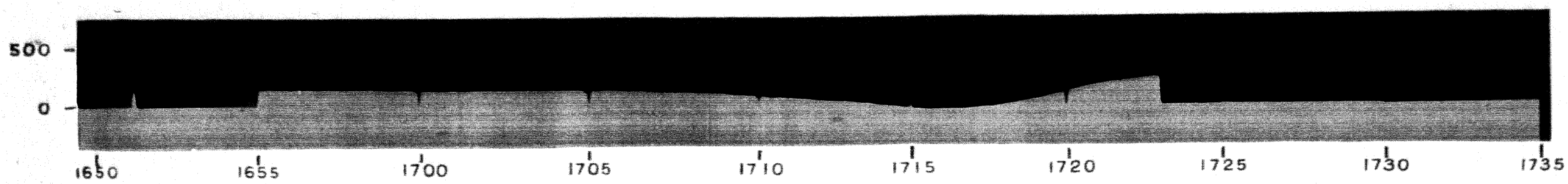


Fig. 2, May 12, 1917.

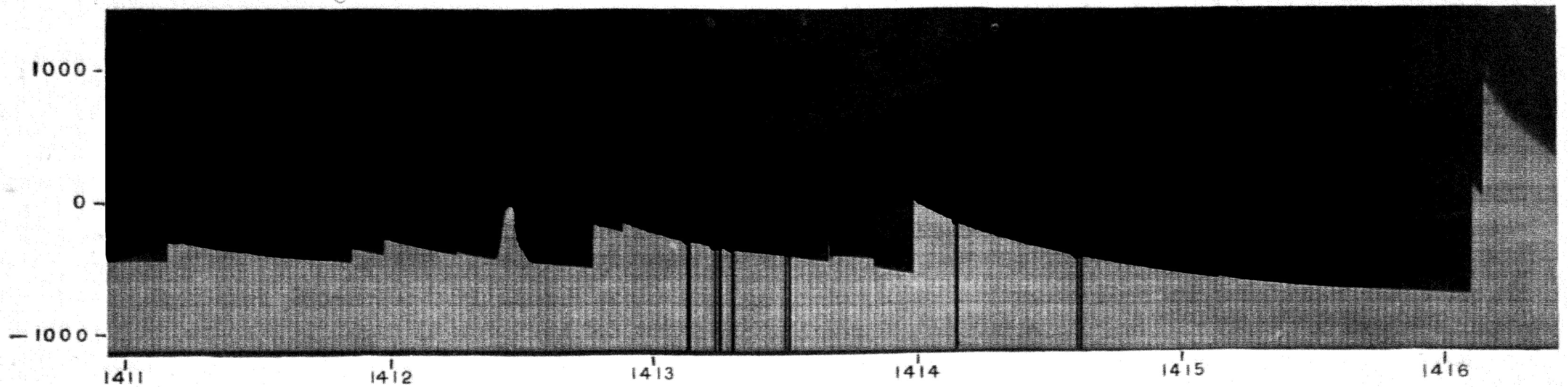


Fig. 3, June 13, 1917.

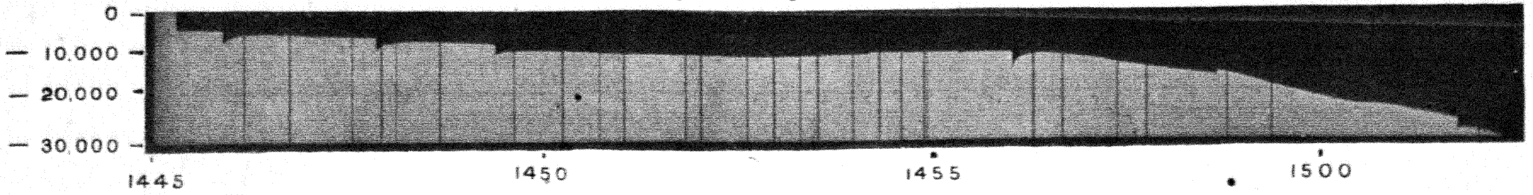


Fig. 4. August 9, 1917.

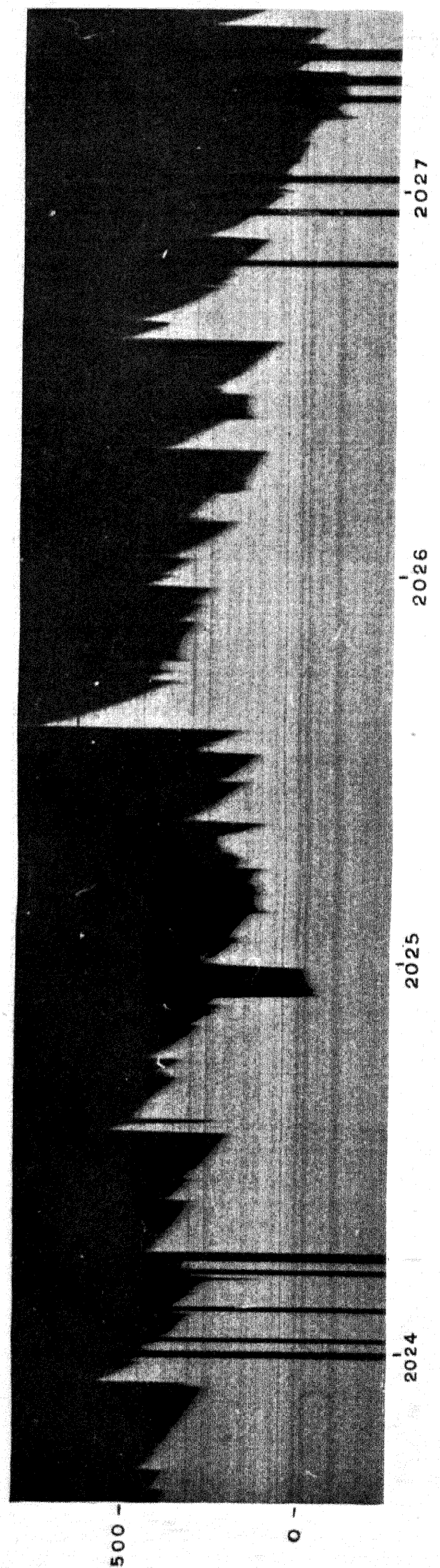


Fig. 5, June 17, 1917

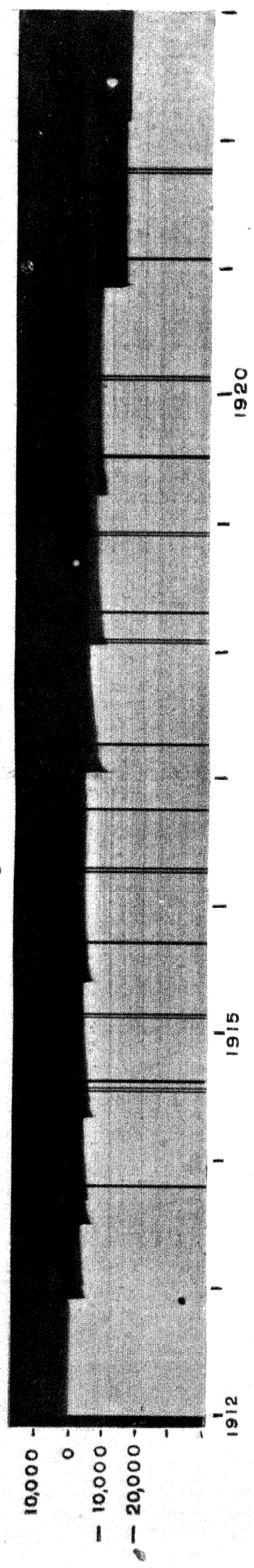


Fig. 6, June 16, 1917.

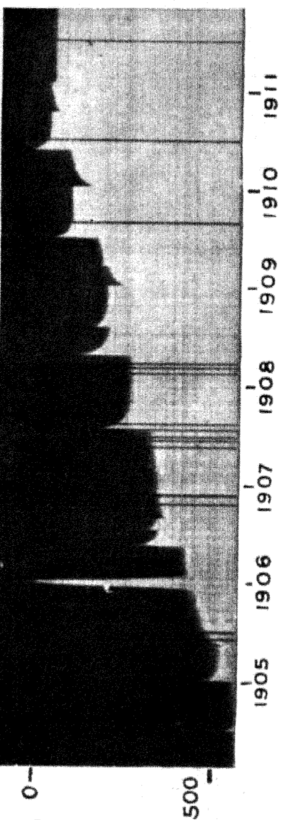


Fig. 8, May 29, 1917.

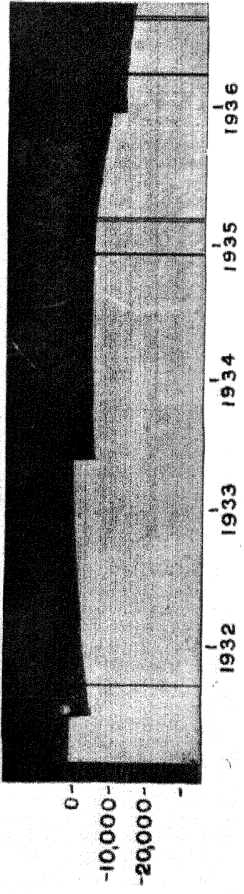


Fig. 7, June 16, 1917

Fig. 9, May 29, 1917.

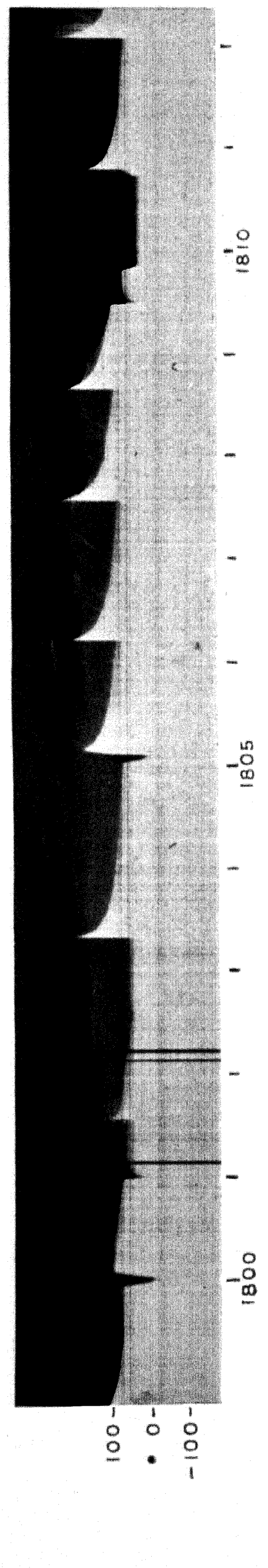


Fig. 10, August 15, 1917.

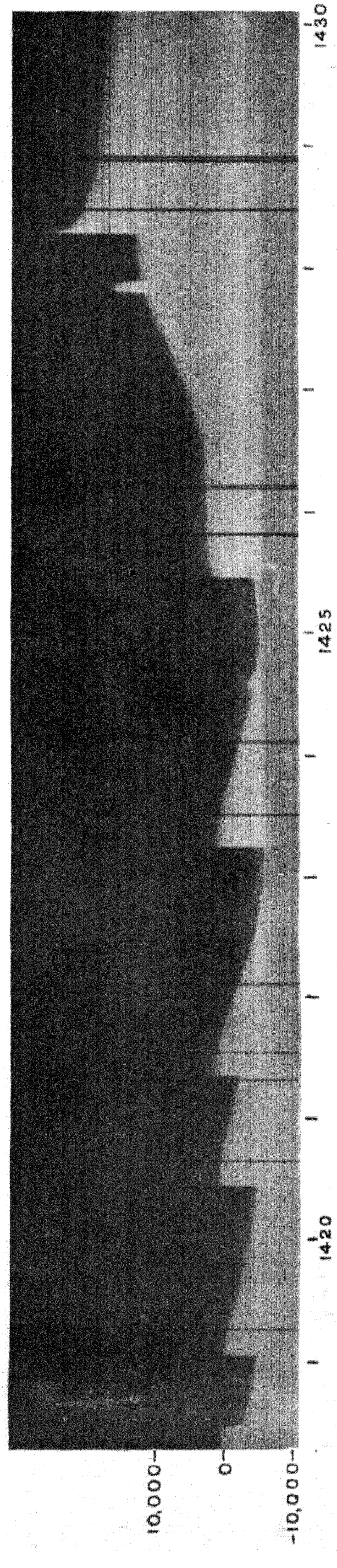


Fig. 11, June 12, 1917.

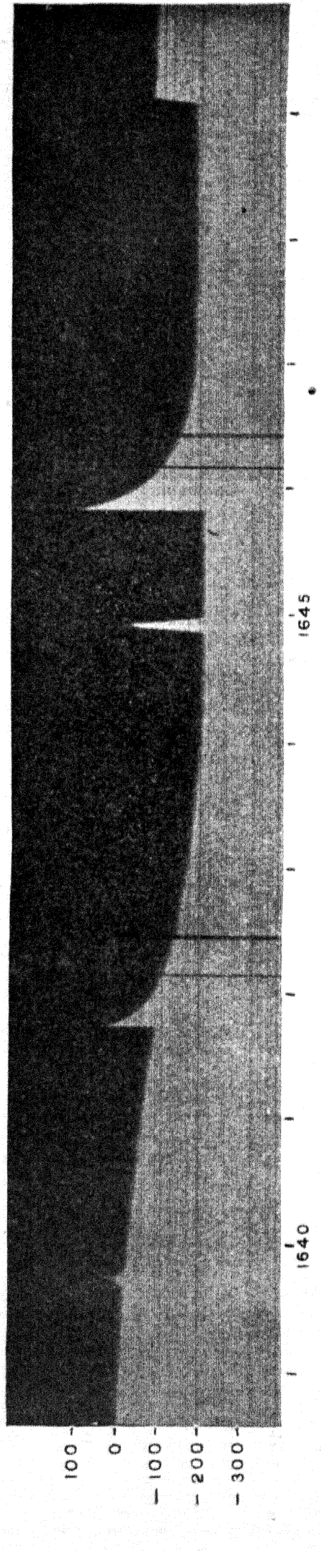


Fig. 12, May 29, 1917

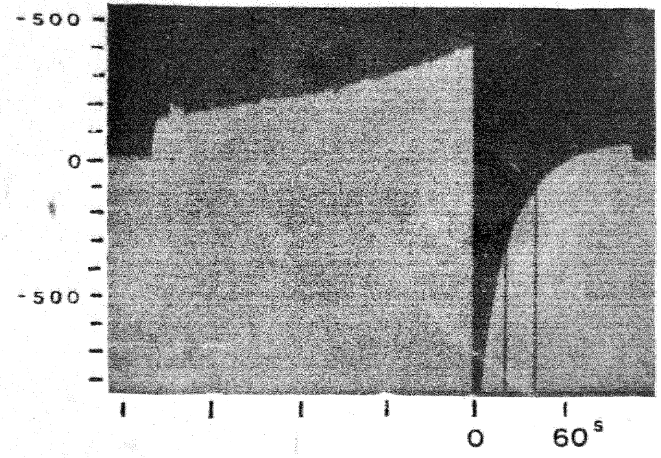


Fig 13, Aug. 15, 1917

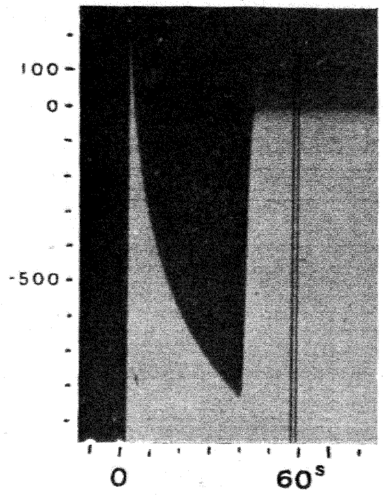


Fig. 14, June 13, 1917

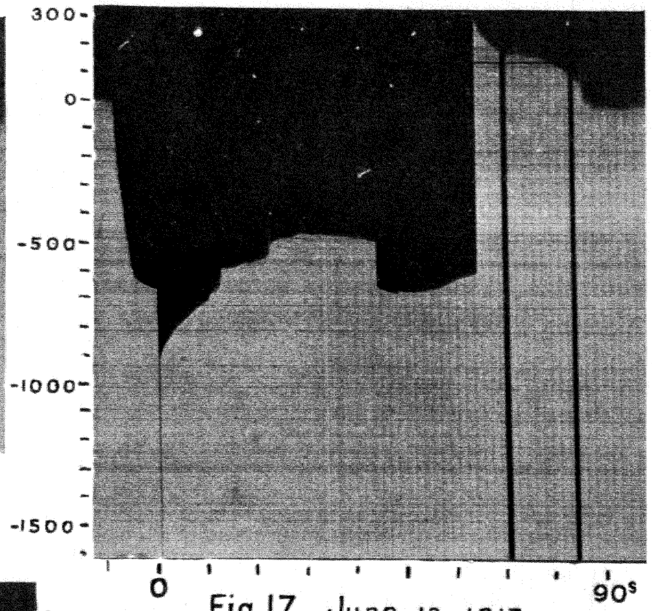


Fig. 15, August 9, 1917

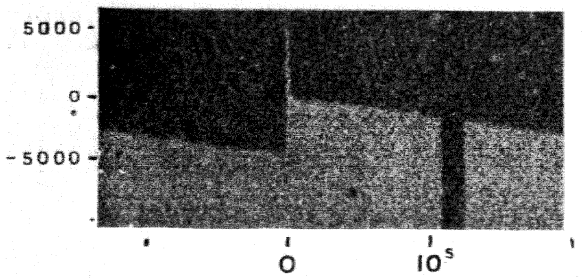


Fig. 16, June 13, 1917

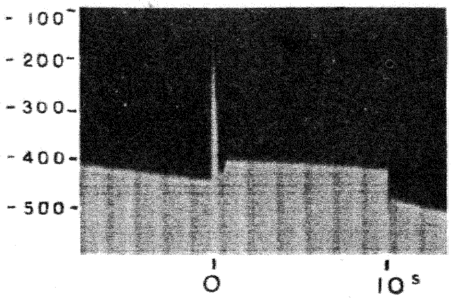


Fig. 17, June 13, 1917

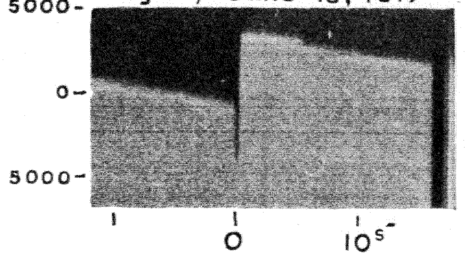


Fig. 18, June 16, 1917

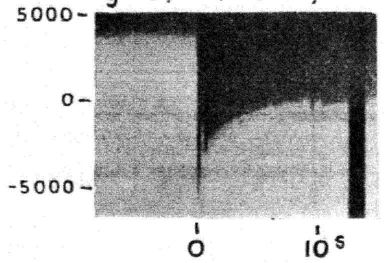


Fig. 19, June 16, 1917

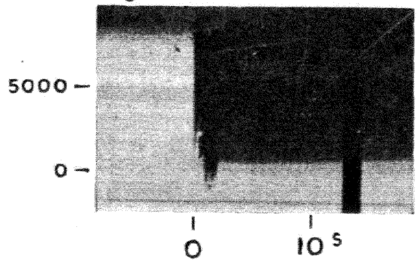
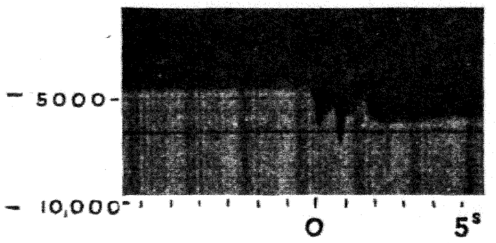


Fig. 20, June 16, 1917



of the whole tube.* If the thickness of the stratiform cloud were small, intense fields might result within the cloud and discharges might even occur; each flash would discharge only a small area of the cloud, of dimensions comparable with the thickness of the cloud.

The characteristic striated and mammatiform appearances frequently observed on the lower surfaces of stratiform clouds associated with thunderstorms may be due to intense electric fields produced as above suggested, the electrical attraction between the upper and lower charges giving rise to convection currents.

If the ionization above and below a stratiform cloud in the field of a primary thunder-cloud is unequal, the cloud will acquire unequal upper and lower charges and thus carry a resultant charge. For example, a stratiform cloud above a cumulo-nimbus cloud will intercept the flow of ions from the upper atmosphere and become charged with electricity of opposite sign to that of the upper pole of the shower-cloud, a steady condition not being reached until a potential difference between the thunder-cloud and the upper atmosphere is concentrated almost entirely in the region below the stratiform cloud. Lightning discharges between the stratiform cloud and the head of the primary thunder-cloud below will be likely to occur.

In the absence of any such cloud above the primary thunder-cloud, the great diminution of the mobility of ions or electrons dragged out of the conducting layers as they penetrate into the denser regions of the atmosphere will have a very similar effect; the concentration of charge will be greatest where the change of conductivity with the height is most rapid. We may in fact, as suggested in Section VIII., consider that the conditions are much the same as if a conducting protuberance were drawn out from the conducting layer towards the summit of the thunder-cloud. It does not seem unlikely that discharges may sometimes occur between this protuberance and the top of the thunder-cloud. In a previous paper some evidence was obtained suggesting the occurrence of discharges of very great vertical length; possibly these may have been of the type we have been considering.

---

* This action of a layer of cloud, in particular of a ground fog, in increasing the vertical electric field within it has long been recognised in the case of the potential gradient of fine weather. ELSTER and GEITEL, 'Meteor. Zeitschr.,' 17, p. 230, 1900; GEITEL, 'Physikal. Zeitschr.,' 17, p. 455, 1916.

# IV. *Researches on the Elastic Properties and the Plastic Extension of Metals.*

*By* W. E. DALBY, *F.R.S., Professor of Engineering at the City and Guilds (Engineering) College of the Imperial College of Science and Technology.*

Original Paper received January 28,—Revised form March 24, 1920.

## § 1. *Preliminary.*

ON March 7, 1912, I described an instrument which gives photographically a load-extension diagram of a metal test piece during the process of stretching it to fracture.

On February 13, 1913, I described further experiments with the instrument.* A diagram was shown which was taken from a test piece broken in ten seconds. It is safe to say that up to that time no apparatus existed which would give a complete record of the load-extension relation during such a quick break.

I have since that time arranged the apparatus to record at even a quicker rate. Fig. 1 shows the record of a break done in 2·15 seconds. The test piece measured one inch between the shoulders. The line of the record is in dashes. These dashes fix the time scale of the diagram. Centre to centre of a pair of dashes corresponds to

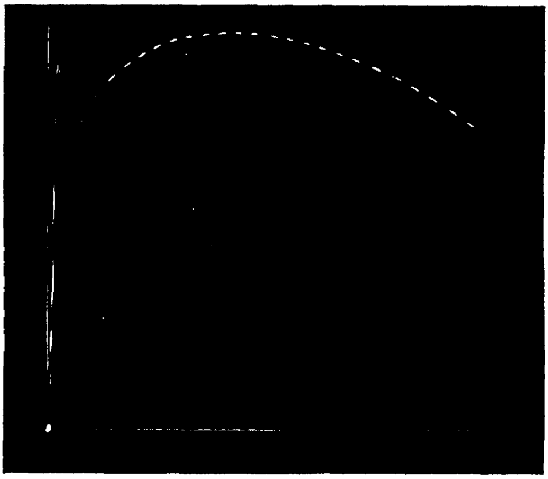


Fig. 1 (mild steel)

---

* Additional results are given in 'Transactions of the Institute of Naval Architects,' March 29, 1912, 'Institute of Metals,' May Lecture, 1917, vol XVIII., No. 2.

[Published September 29, 1920

$\frac{1}{14}$ of a second. This time calibration is obtained by placing an interrupter in the lamp circuit. Referring to the diagram it will be seen that the yield load was reached in about $\frac{7}{14}$ seconds.

In my 1913 paper I included a diagram taken with an instrument which multiplied the extension of the gauge length 150 times so that the elastic part of the curve appeared on a scale which enabled its shape to be studied and which enabled the limit of proportionality to be identified when such a limit existed.

In my method of taking these diagrams the test piece is stretched without pause in the loading and the spot of light follows without break of continuity every phase of the relation between load and extension. Sudden slips of the crystals are duly recorded.

In the usual method the load is applied in steps, pausing at each step to observe the extension, so that the piece gets a rest under steady load during the time occupied in making the observation of extension. The load-extension curve is thus defined by a definite number of points only and peculiarities of form between these points are missed.

I have from time to time continued these elastic researches, and the following paper records some of the results obtained with what I call the *Optical Recorder of Load and Elastic Extension.*

## § 2. *The Test Piece.*

In these researches the gauge length is defined by flanges turned on the test piece itself. The ends of the arms of the extensometer rest on these flanges. The dimensions of the standard form of test piece used in these researches are shown in fig. 2. A shorter gauge length was used for the more ductile metals, but all experiments on the iron and steels were done on a 5-inch gauge length.

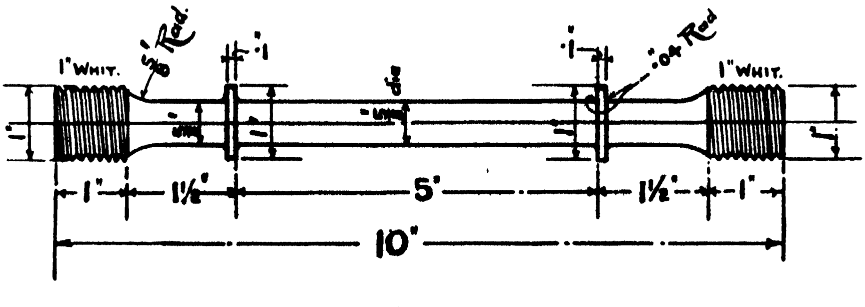


Fig. 2.

I was lead to adopt this form by the many difficulties encountered when pointed screws have to be driven into the test piece to define the gauge length.

These screws cannot be driven properly into hard material like the alloy steels which have to be tested nowadays, and in soft material like copper the primitive centre dot in which the points rest, elongates as stretching proceeds and the points slip.

The modulus of elasticity, E, found from flanged test pieces agrees with the values found from plain bars. The flanges therefore have negligible influence on the elastic extension of the gauge length. They restrict the plastic extension slightly. In mild steel the total extension is about 3 per cent. less when found from a flanged test piece than it would be if found from a plain bar.

Dr. COKER has kindly examined the distribution of stress produced by a flange in a xylonite test piece made to the dimensions of fig. 2. When the xylonite test piece is stretched the colours show that there is no stress in the flange itself and there is a slight but symmetrical modification of the stress distribution at its root. This means that as stretching proceeds the flange is not distorted, and therefore the distance between the flanges is a true measure of the extension of the primitive gauge length which they define.

### § 3. *The Elasticity of Materials and a Typical Load Elastic Extension Diagram of Mild Steel.*

The elasticity of a material means in a general sense its power of returning to its primitive form after loading has been applied and removed.

The recovery may be partial or complete.

The power of complete recovery is lost when the stress produced by the loading has once passed beyond a certain limiting value peculiar to the material.

Below this limiting stress the extension of a steel test piece is proportional to the load.

Above this limiting stress the extension increases at a greater rate than the load.

The limit is therefore called THE LIMIT OF PROPORTIONALITY.

The power of recovery may thus be distinguished into the power of complete recovery possessed and retained only so long as the stress in the steel has never once exceeded the limit of proportionality: and the power of partial recovery peculiar to the state into which the metal passes directly it has once been loaded beyond the limit of proportionality.

Provided that the material has never been loaded beyond its limit of proportionality the material may be said to be in a state of *perfect elasticity*, because it possesses the power of complete recovery of form after removal of load; alternatively it may be said to be in a state of *proportional elasticity* because its extension is found to be proportional to the load.

The one term includes the other. If it is found to extend proportionally to the load its recovery is perfect after removal of load.

No metal is, however, quite perfect in its recovery, but the term perfect used in the sense defined above is convenient and substantially expresses the experimental results.

A diagram recording the elastic extension of mild steel is seen in fig. 3.

This steel contains 0·156 per cent. of carbon. The extension scale of the diagram is defined by the distance between the two vertical lines seen in the diagram. This distance represents an extension of 0·01 inch.

Proportionality between load and extension ceases at about 4·5 tons corresponding to a stress of 14·67 tons per sq. inch. Yield occurs at 6·7 tons which corresponds to 21·85 tons per sq. inch. The load drops away from yield to about 5·5 tons giving a stress of 18 tons per sq. inch. These stress are reckoned on the original area of the cross-section of the test piece.

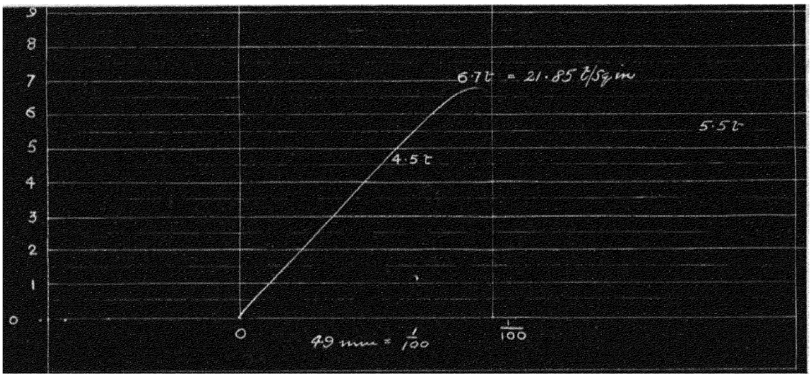


Fig. 3 (mild steel).

The slope of the line from the origin to the limit of proportionality defines E, the modulus of elasticity. From the diagram its value is 13,300 tons per sq. inch.

## § 4. *Restoration of Perfect Elasticity after Overstrain.*

The term "OVERSTRAIN" means that a metal has been loaded beyond its limit of proportionality.

If the load is removed after a test piece has been strained beyond the limit of proportionality and then the piece is immediately re-tested, the record shows a curved line.

It has no range of proportionality and no modulus of elasticity which can be identified with E.

The material still possesses elasticity because it shrinks as the load is removed, but the elasticity is imperfect in the sense that change of length is no longer proportional to change of load.

But, if the metal is iron or mild steel, proportional elasticity is slowly recovered with time; and this change from unproportional to proportional elasticity or from imperfect to perfect elasticity is accelerated by boiling.

In fact overstrained iron or mild steel is restored to its perfect or proportional elastic state with remarkable rapidity by mere boiling. This point has been established by Sir ALFRED EWING.*

I have found, however, that overstrained high carbon steels and the alloy steels do not recover proportional elasticity either by resting or by boiling.

The elastic line of a 3 per cent. nickel steel is seen in fig. 4. It is lettered A.

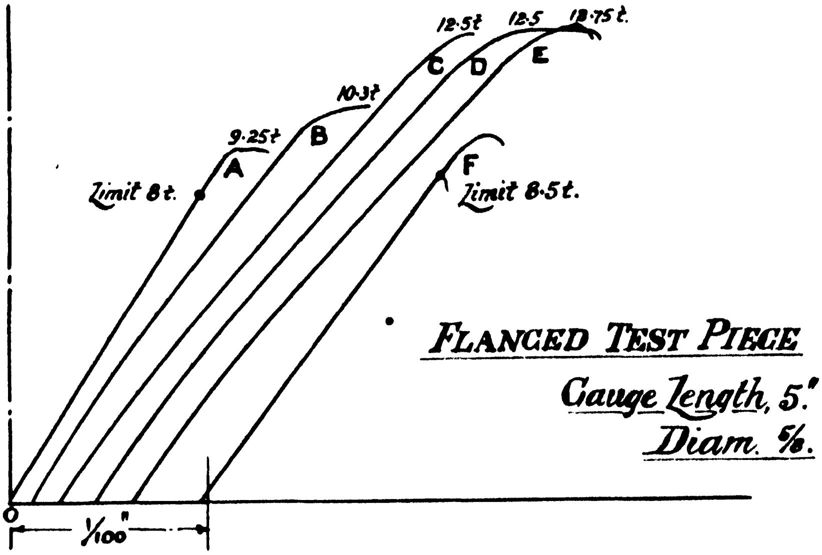


Fig. 4 (nickel steel).

The limit of proportionality is reached at 8 tons, 26 tons per sq. inch, and the yield at 9·25 tons, 30 tons per sq. inch.

The piece was stretched 2 per cent. and then the diagram, line B, was taken. Proportional Elasticity has disappeared.

Curve C is the record after a 6 per cent. stretch.

Curve D is a repetition test after turning the bar to a slightly reduced diameter. The interval of time between C and D is 24 hours. No restoration of elasticity has taken place. It has been established by other experiments that a lapse of many months has no effect in restoring the proportional elasticity.

The piece was then boiled for 1 hour, and curve E shows that elasticity has not been restored.

Finally, the piece was heated to 550° C. in a muffle furnace for about half an hour and was then allowed to cool down with the furnace. Line F, taken immediately after this treatment, shows perfect recovery of proportional elasticity and a slight raising of the limit and the yield point.

I have confirmed these results by other experiments on nickel steel test pieces and on high carbon steel test pieces.

---

* 'Phil. Trans. Roy. Soc.,' 1899.

### § 5. *Looping after the Elastic Limit of Proportionality has been Passed.*

The recorder is fitted with a microscope so that the process of stretching can be watched as the experiment proceeds, and the loading, which is produced by hydraulic pressure, stopped at any moment. This arrangement enables interesting records to be taken, because after the test piece has been stretched an assigned amount, the load can be let off and then immediately re-applied, so that stretching continues through a second interval and so on.

Such a record is seen in fig. 5. The material is nickel steel. It will be seen that the removal and the re-application of the load compels the spot of light to trace a loop. The area of the loop represents the internal work done during the process.

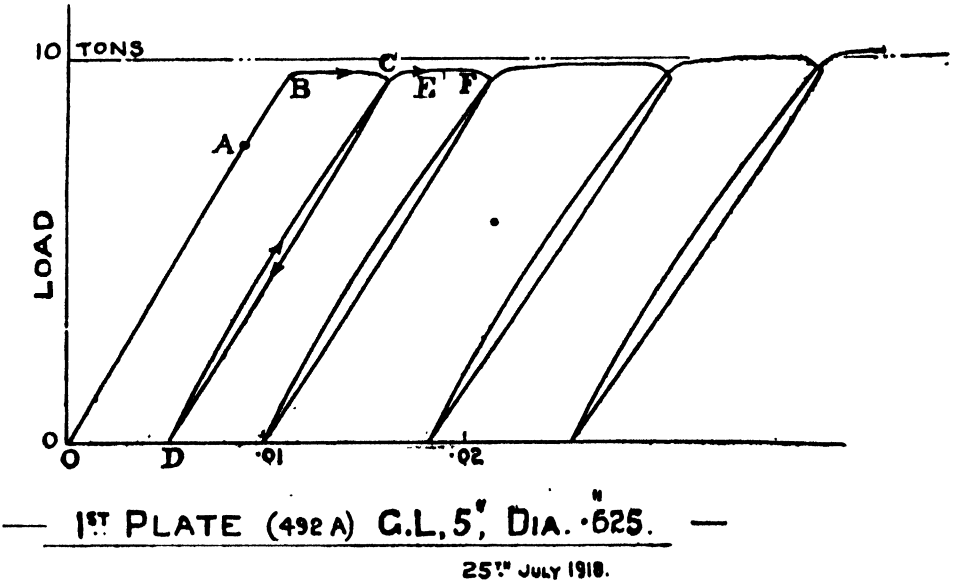


Fig. 5 (nickel steel).

Following the path of the spot it starts from the origin O and describes the elastic line OA, passes the limit of proportionality at A, and then curves away to the yield point B, and on to C. At C the loading is stopped, the hydraulic pressure is relieved by opening the exhaust valve, and the spot travels down the curved path CD as the load falls to zero. The exhaust valve is then closed and the pressure valve is opened and the process is repeated through the path EF and so on.

When the steel test piece has been stretched beyond its limit of proportionality, for example to C, fig. 5, the total extension is made up of two parts, namely :—

(1) the proportional elastic extension up to the limit of elasticity, for example up to A, fig. 5;

(2) the plastic extension after the limit has been passed.

When the load is removed the steel shrinks unproportionally, for example from C to D, by an amount approximately equal to its previous proportional elastic extension and then stops at a dimension greater than its primitive dimension by an amount approximately equal to its plastic extension.

The increase of size measured after the removal of the load is called the Permanent Set. OD is the permanent set produced by the first stretching of the test piece to C. From the diagram the permanent set measures 0·0052 inch and the unproportional elastic recovery measures 0·011 inch.

When the load is re-applied the spot of light moves from D to E along a curved path. The extension is no longer proportional to the load. The metal is in a different elastic state. Stretching beyond the limit of proportionality has robbed the metal of its power of proportional extension and of perfect recovery after removal of load. It may be said to be in a state of *imperfect elasticity*, or alternatively it may be described as in a state of *unproportional elasticity*.

The imperfect state is disclosed by the loop formed by the removal and re-application of the load.

The diagram shows four loops. Each loop is slightly larger than the loop preceding it.

The four loops shown were all recorded on a half plate inserted in the camera. A succession of plates was taken and the last plate is shown in fig. 6.

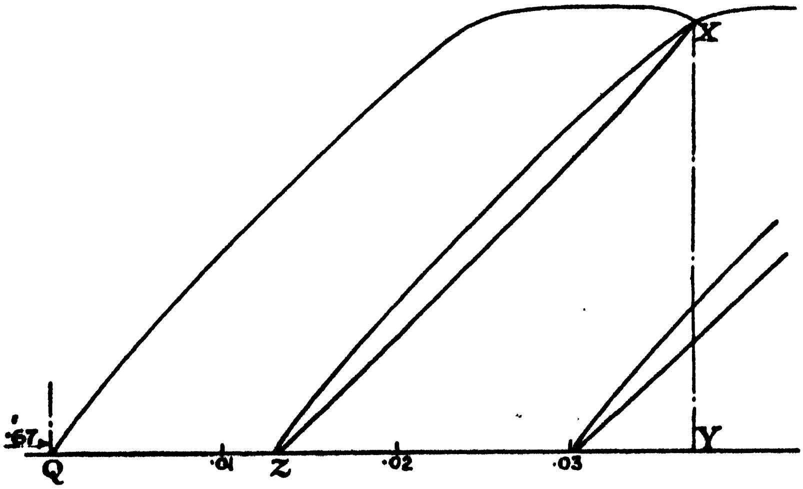


Fig. 6 (overstrained nickel steel).

This last plate shows that unproportional elastic shrinkage occurs right up to the load at which local contraction begins. The last loop is just seen on the plate.

The last line seen curving up from the origin Q is the typical curve of overstrained material. The primitive gauge length of 5 inches had been stretched to 5·67 inches before the last plate, fig. 6, was taken.

The permanent set QZ measured from the new origin Q is 0·0127 inch. The total permanent set is thus 0·67 + 0·0127 inch = 0·6827 inch. The unproportional elastic recovery is ZY = 0·0244 inch.

### § 6. *Looping a General Property of Metals.*

Records of looped diagrams are shown in the following figures. The load scale is varied to bring out the shape of the loops.

The extension scale is substantially, $1\frac{1}{2}$ inches measured horizontally on the diagram represents $\frac{1}{100}$ inch extension of the gauge length. The diagrams are reproduced as taken, the object being to compare the elastic line and the loop formations.

*Staffordshire Iron.*—Fig. 7. The limit of proportionality is reached at 3·5 tons; 11·4 tons per sq. inch. The yield is reached at 5·1 tons; 16 tons per sq. inch. The load then drops to about 4·7 tons. The loop area is not large, but the area increases progressively.

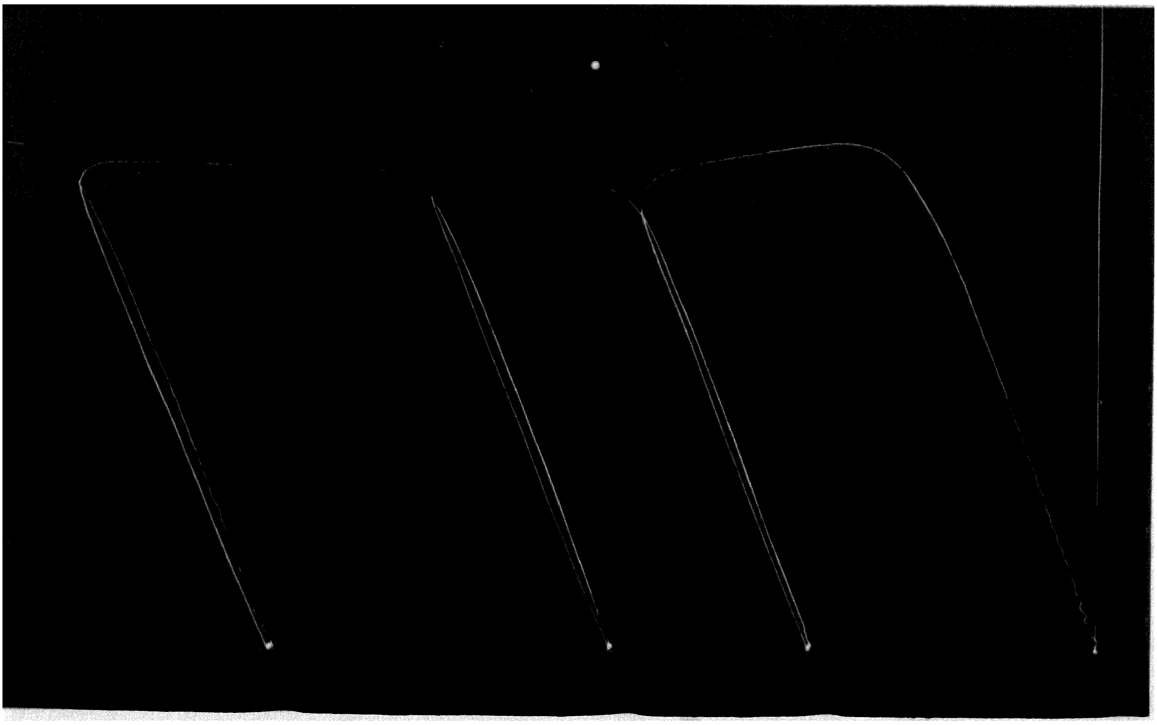


Fig. 7 (iron).

*Steel.*—Carbon 0·8 per cent., fig. 8. The limit of proportionality is reached at about 8 tons; the yield at 8·85 tons. There is a slight drop at the yield load. The area of the first loop is many times larger than the area of the first loop of the previous record, and this area increases rapidly in the succession of loops.

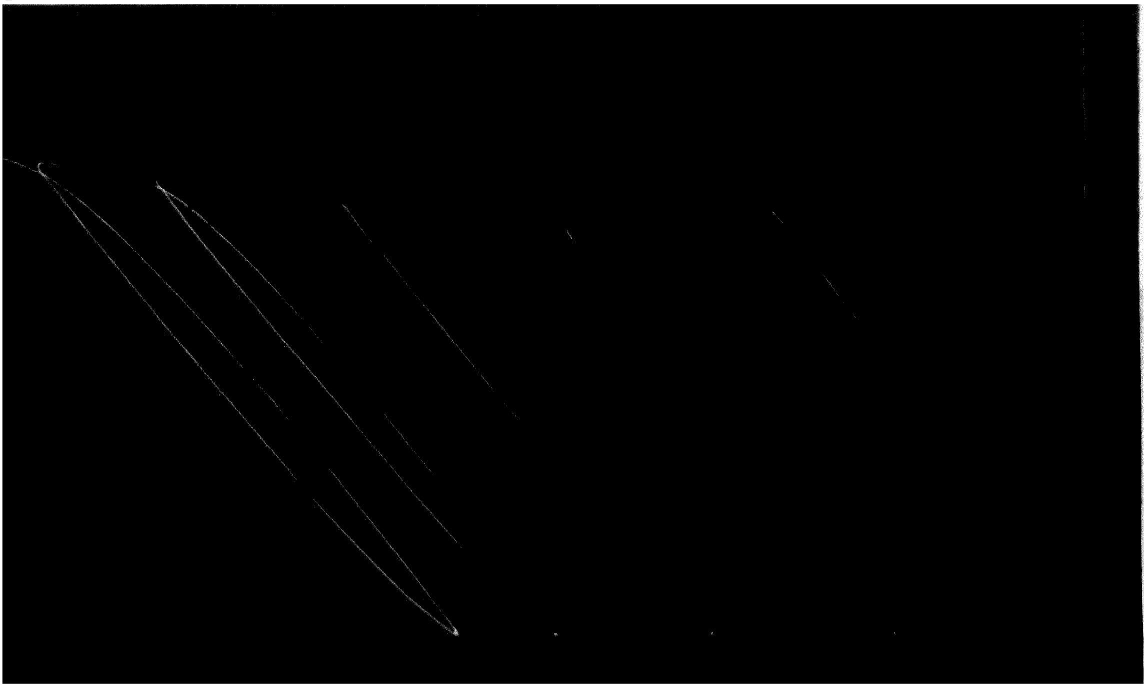


Fig. 8 (0·8 carbon steel).

*Steel.*—Carbon 0·8 per cent., fig. 9. This diagram is introduced because it is taken from a test piece cut from a bar delivered from the works as steel of the same kind and quality as that from which the previous diagram was taken.

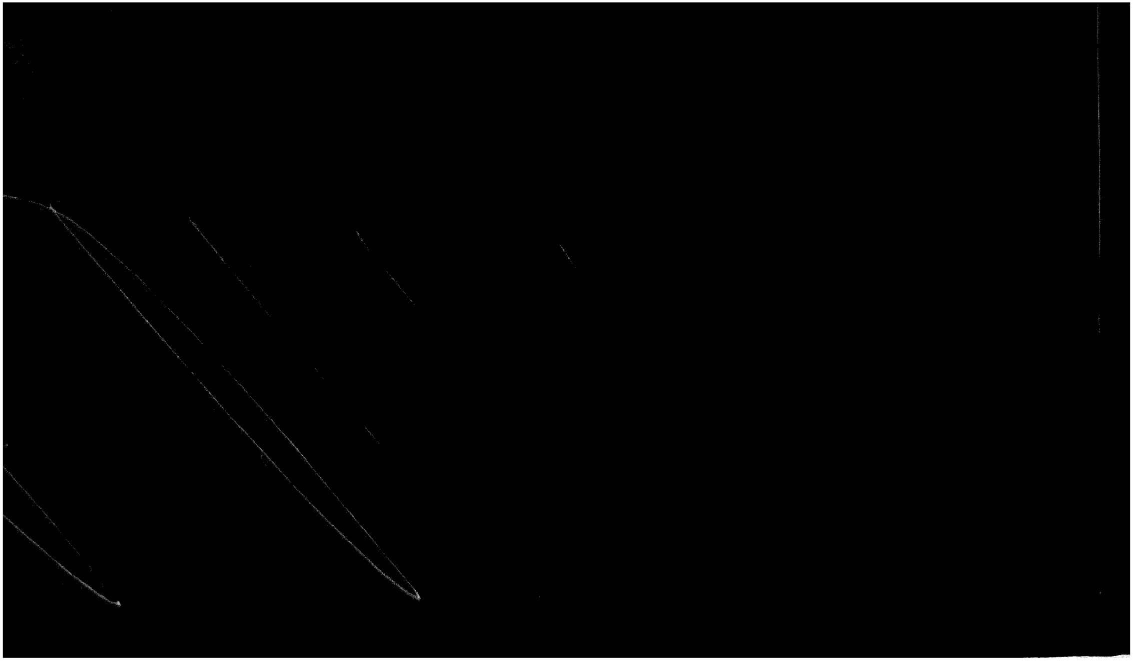


Fig. 9 (0·8 carbon steel).

The limit of proportionality occurs now at about 3 tons and there is no definite yield point. The loop areas and their rates of increase are about equal in the two plates.

The explanation of the difference in quality sho[illegible]omparing the two diagrams may be found in the fact that the test piece of fig. 8 was cut from a bar delivered before the war. The test piece of fig. 9 was cut from a bar delivered towards the end of the war. There has clearly been some change in the manufacturing process.

*Nickel Chrome Steel.*—Fig. 10. The ultimate strength of this steel found from a bar 1 inch diameter is 54 tons per sq. inch, with an extension of 14 per cent. on 8 inches and 55 per cent. reduction of area. The limit of proportionality is at a load of 10 tons on the standard test piece 0·625 inch diameter, corresponding to 32·5 tons per sq. inch. Yield sets in at 11 tons, that is, 36 tons per sq. inch.

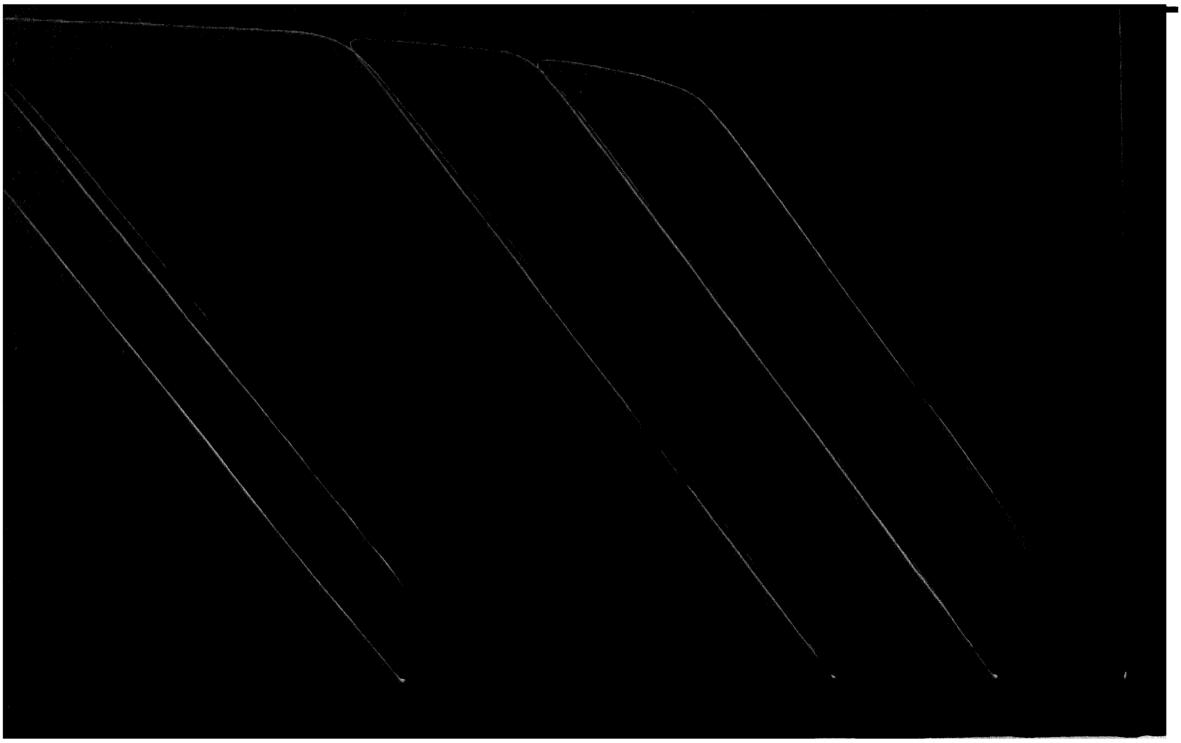


Fig. 10 (nickel chrome steel).

The first loop of the diagram is small, but the area increases rapidly, as will be seen from the three loops visible in the record.

*Nickel Steel.*—Carbon 0·33 per cent., Ni 3·52 per cent.—Fig. 11. The ultimate strength of this material is about 48 tons per sq. inch, with an elongation of 20 per cent. on 5 inches and a reduction of area of 47 per cent. Limit of proportionality occurs at about 30 tons per sq. inch and yield at 32 tons per sq. inch. The limit of proportionality here approaches quite near to the yield point.

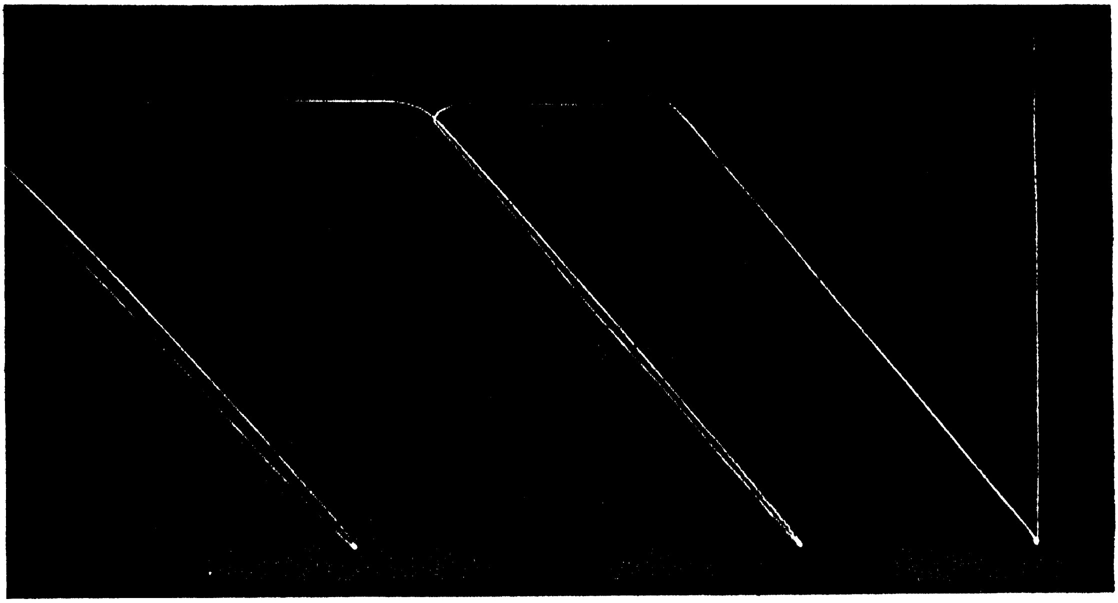


Fig. 11 (nickel steel).

*Zinc.*—Fig. 12. Diameter of test piece 0·8 inch. Gauge length 5 inches. The test piece was turned from a zinc rod. There is no proportional elastic line. Curvature

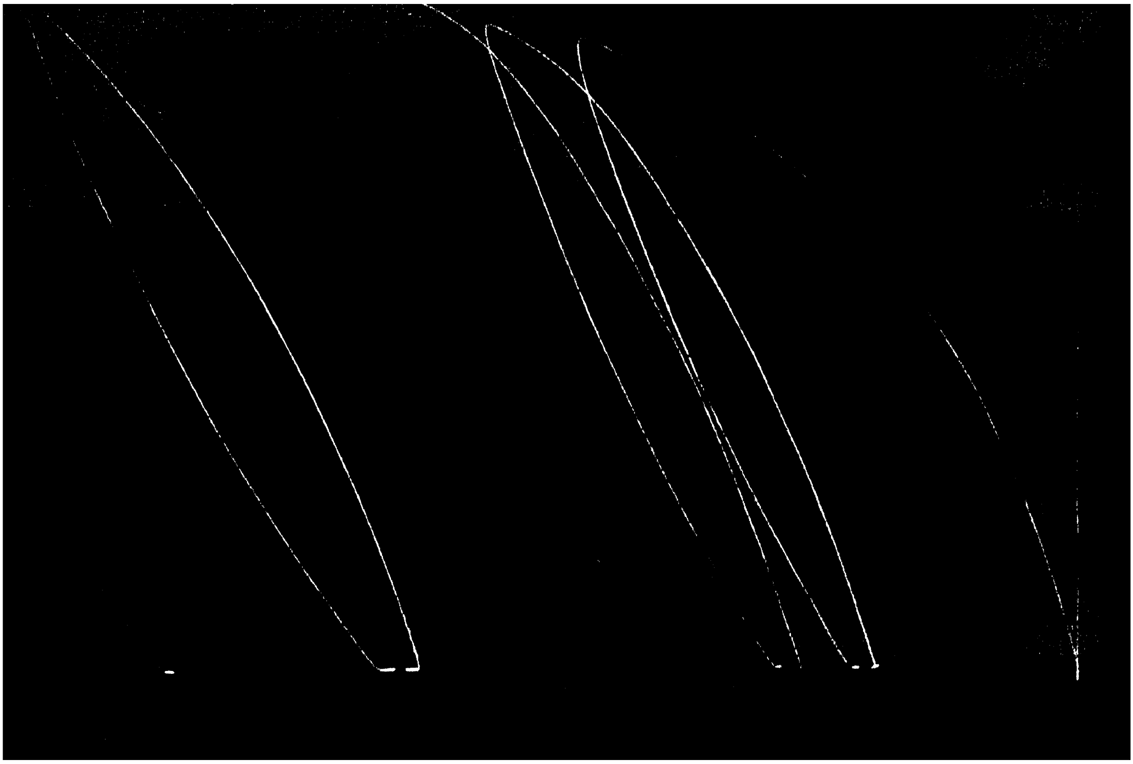


Fig. 12 (zinc).

begins at the origin, so that the extension is increasing at a greater rate than the load from the commencement of loading. There is no definite yield point.

The most interesting result to notice is that the test piece goes on shrinking in length after the load has been removed. It is shrinking under the action of its own internal molecular forces because it is entirely free from external load.

The shrinking at no load is indicated by the flat bottom of the loop. A dwell of 1½ minutes was made in the experiment after the removal of the load and before the re-application of the load. All the perceptible shrinking at no load takes place within this time interval.

At the third loop after the load was removed the light was shut off and flashed on at intervals of $\frac{1}{10}$ seconds to get some idea of the rate of shrinking.

*Tin.*—Fig. 13. Diameter of test piece 0·8 inch. Gauge length 5 inches. This test piece was turned from a bar of tin. It exhibits properties similar to zin on a smaller scale. There is shrinking continuing for about 1 minute after th load has been removed, and there is the same absence of a proportional elastic line.

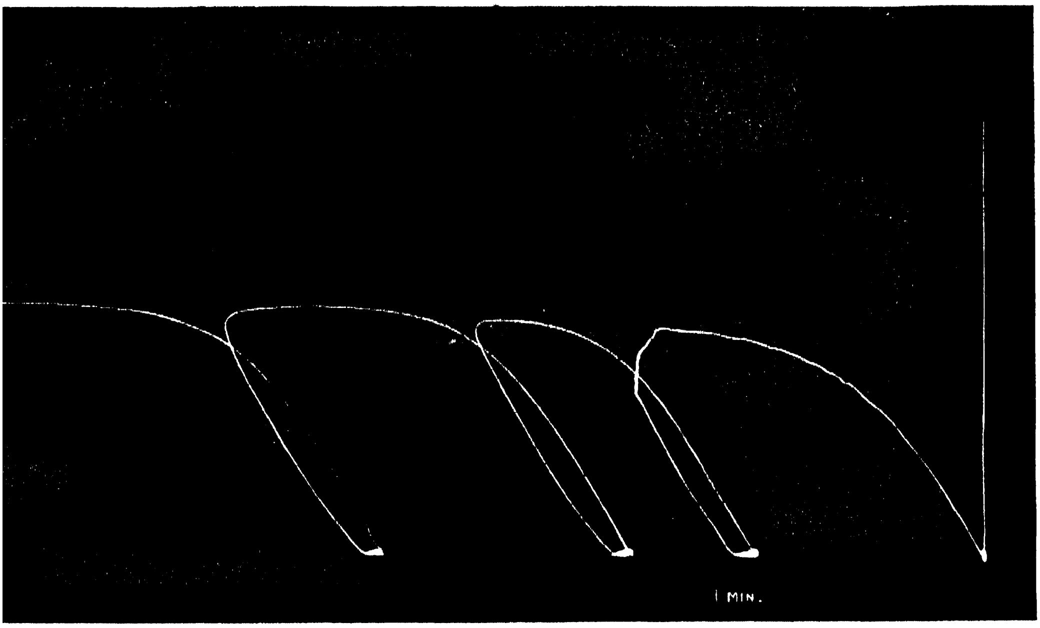


Fig. 13 (tin).

*Copper.*—Pure and free from arsenic. Fig. 14. Diameter of test piece 0·8 inch Gauge length 5 inches. There is no marked limit of proportionality and no yiel point in this material. The noteworthy feature of the record is the small rate increase of loop area. This small rate of increase of loop area is a commo characteristic of all the copper samples which I have tested. It may be that this rate of increase is identified with the quality of toughness.

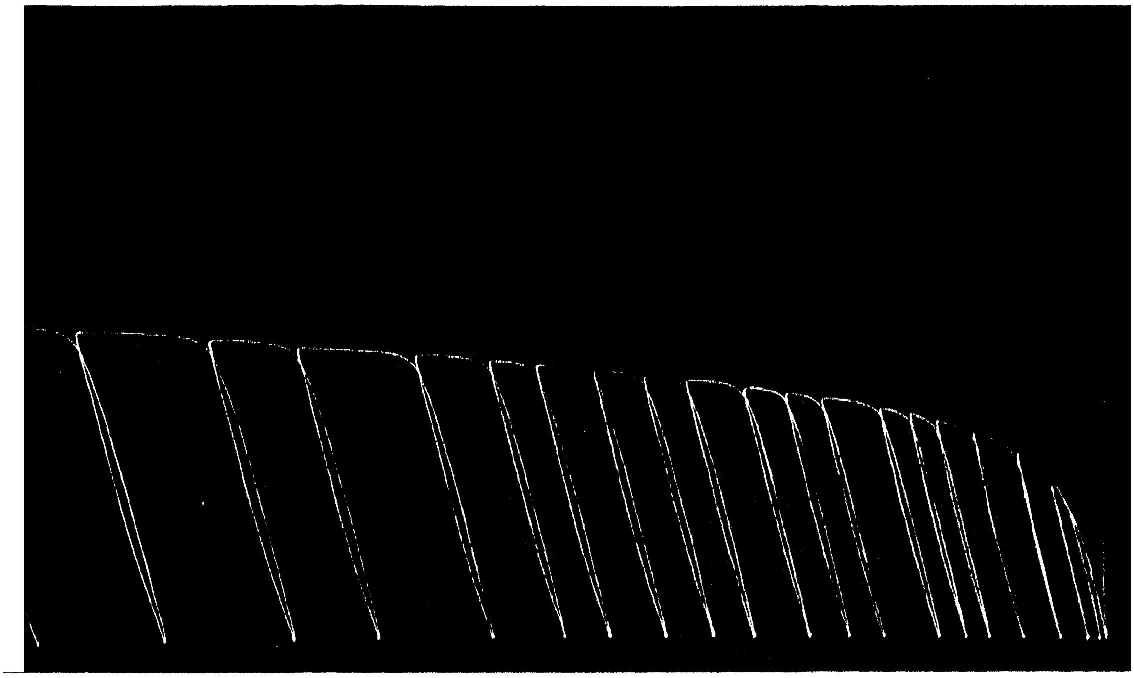


Fig. 14 (electrolytic copper).

*Copper.*—Fig. 15. 99·4 per cent. copper. Arsenic present, and by difference estimated at 0·4 per cent. Diameter of test piece 0·8 inch. Gauge length 5 inches.

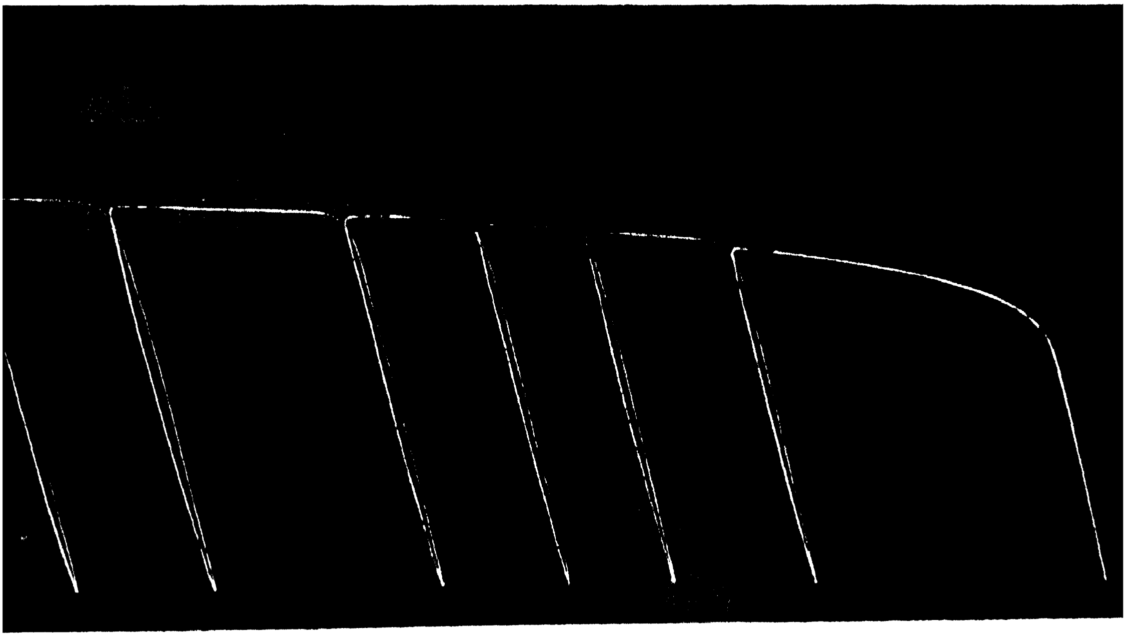


Fig. 15 (arsenical copper).

The effect of the arsenic is remarkable. It gives to the copper an elastic line with a distinguishable limit of proportionality of 1·4 tons; 2·8 tons per sq. inch.

There is no definite yield point. The loops are small and the rate of increase of loop area is small.

The elastic line from the origin to 1·1 tons is thicker than the continuation of the line. This thickening is brought about by a removal and a re-application of the load. The spot of light travelled three times up and down this piece of the diagram, indicating that the elastic line, within the limits of this load, is permanent.

*Brass.*—Fig. 16. Composition 60 per cent. copper, 40 per cent. zinc, with traces of tin and other impurities. The ultimate strength of the material is 32·6 tons per sq. inch. There is a marked limit of proportionality at 2¼ tons; 7·33 tons per sq. inch. Yield follows gradually. There is no contraction after the load has been removed, although the material contains so much zinc.

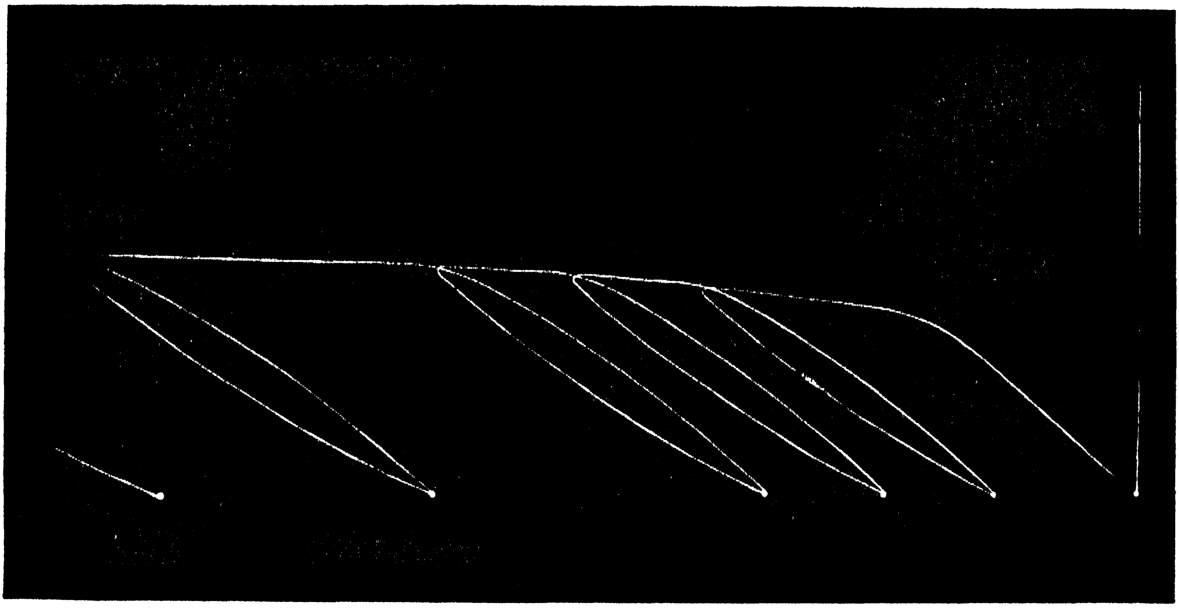


Fig. 16 (brass).

*Phosphor Bronze.*—Fig. 17. The curve in this diagram shows a limit of proportionality at about 2 tons; 6·5 tons per sq. inch; but it is difficult to locate the exact spot at which the line begins to curve away from the primitive straight element.

*Aluminium Alloy.*—Fig. 18. Diameter of test piece 0·625 inch. Gauge length 5 inches. This diagram is remarkable in that the removal and the re-application of the load in the plastic state shows no looping and therefore no hysteresis loss which can be calculated from the loop area. It appears as though the metal continually anneals itself at ordinary temperatures as plastic stretching proceeds. The material appears to be elastic up to a load of 4½ tons. The thick line indicates the removal and the re-application of load before the metal begins to yield plastically.

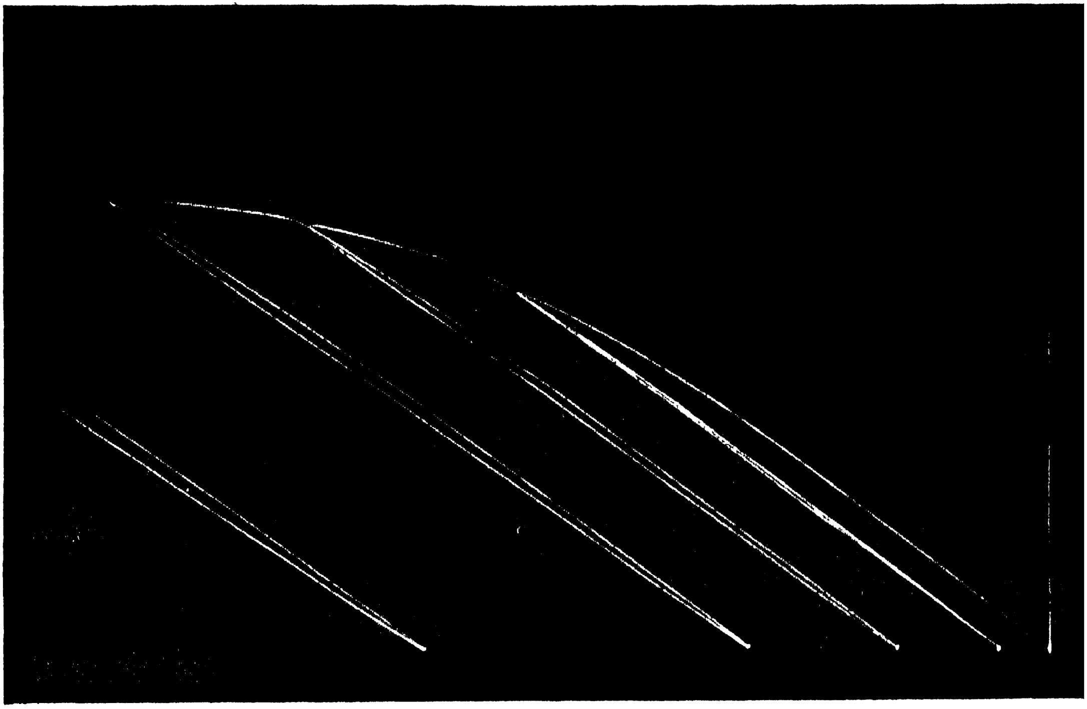


Fig. 17 (gun metal).

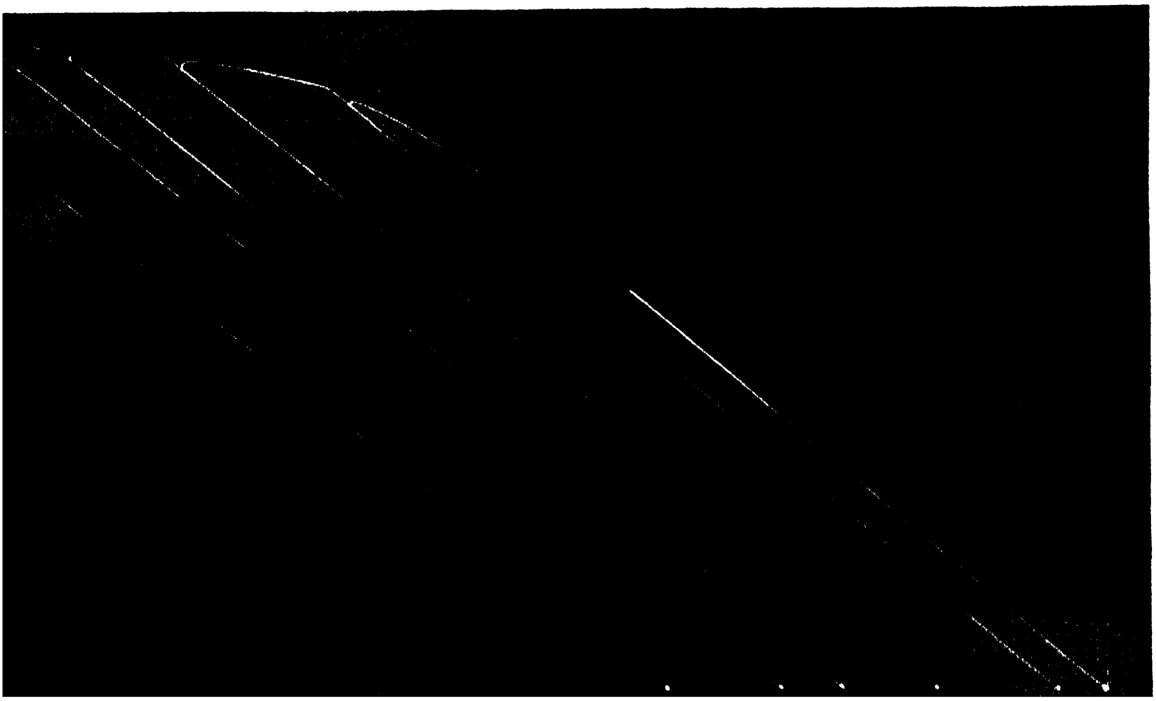


Fig. 18 (aluminium alloy).

### § 7. *Loop Area and Permanent Set.*

The loop area increases in size as the stretching proceeds and the rate of increase differs in different materials.

The question now arises: does the increase in area follow a regular law? The answer is given by the curves on Sheet 1 (folding diagram).

The co-ordinates on Sheet 1 are loop area and permanent set. Curve 1 shows the results obtained from a test piece of 0·8 inch carbon steel, ⅝ inch diameter, with a 5 inch gauge length. The slope of this curve shows the rate of increase of loop area as stretching is continued. The curve ends when local contraction begins. Similar curves are given on Sheet 1 for nickel steel, mild steel, and for iron.

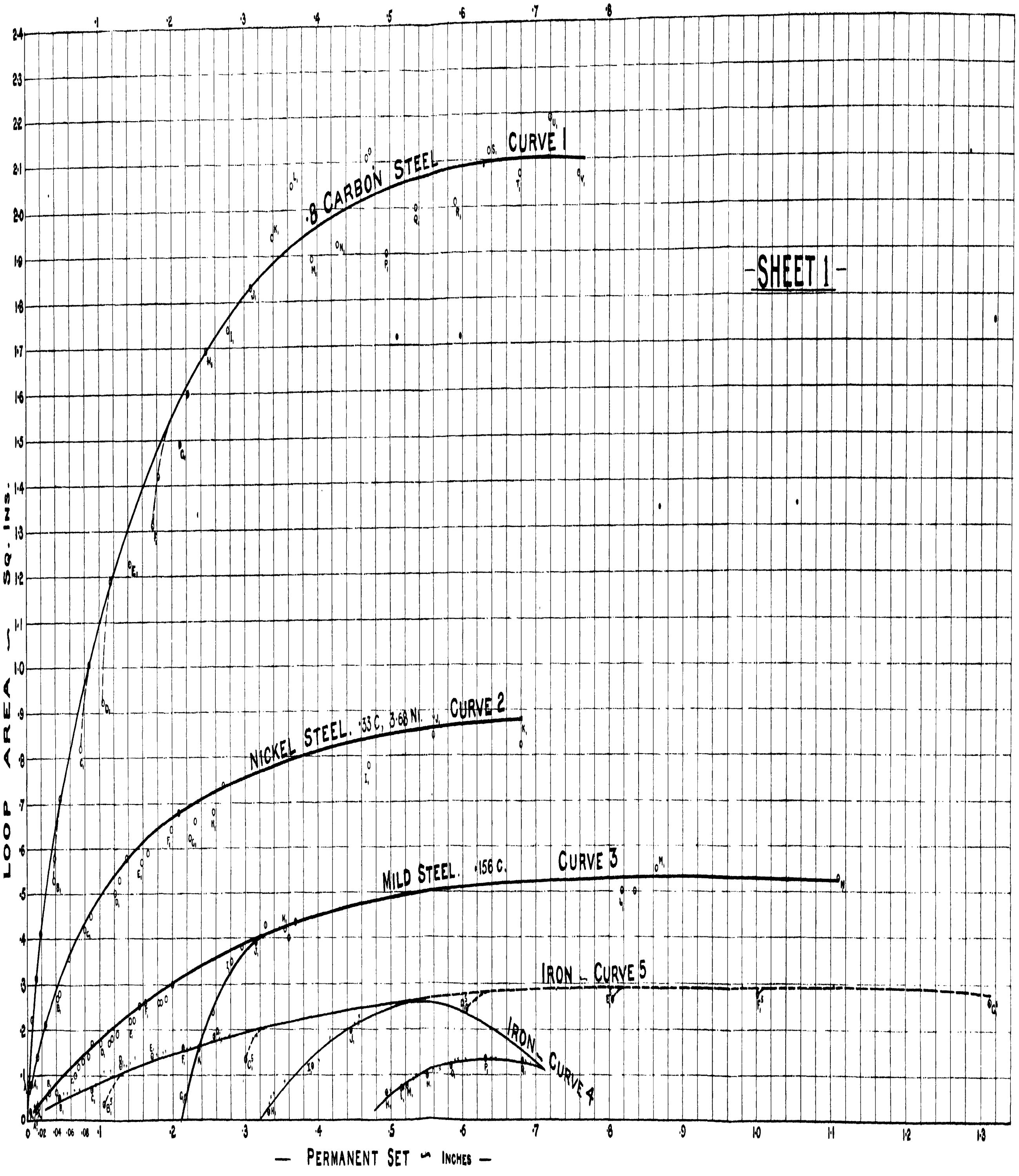
The rate of increase depends upon the time interval between the drawing of the loops and upon the kind of material. In irons and mild steels the influence of time is profound. In the alloy steels tested and in high carbon steels the influence of time is small.

When the stretching of iron or mild steel is resumed after a rest, the loop area, at first small, increases rapidly towards the area the loops would have had if stretching and looping had been continued without resting. Anticipating the detailed description of the curves on Sheet 1, this point may be illustrated by curve 3, same sheet. Plates $A_1$, $B_1$, $C_1$, $D_1$, $E_1$, and $F_1$ were taken consecutively, there being no more time interval between the plates than the few seconds required to change the plates. After the mild steel test piece had been stretched to 0·2 inch it was taken out of the machine and laid aside for 15 days. Plate $G_1$, the first plate taken after the rest, and plates $H_1$, $I_1$, $J_1$ furnish loops of rapidly increasing area until the area is reached on plate $K_1$, corresponding to continuous stretching without rest intervals.

### § 8. *Loop Area and Permanent Set Curve, Curve* 1 *Sheet* 1 (0·8 *per cent. Carbon Steel*).

The detailed consideration of this curve will show how all the curves of the diagram on Sheet 1 have been derived. The capital letters along the top of the procession of loops seen in fig. 19 refer to the sequence of negatives recording the loops taken from a standard test piece of 0·8 inch carbon steel.

Plate A gives the record of the first application of the load to the test piece and its immediate removal and re-application four times. The plate therefore shows the elastic line and the first four loops. A *scale* is placed under the loops so that the *permanent set* of the primitive 5-inch gauge length can be read at any point in the procession of loops. For example, the permanent set at the end of the looping operations recorded on the sequence of plates A, B, C, D, is the distance $0k$ = 0·137 inch.



## — LOOP AREA PLOTTED AGAINST PERMANENT SET

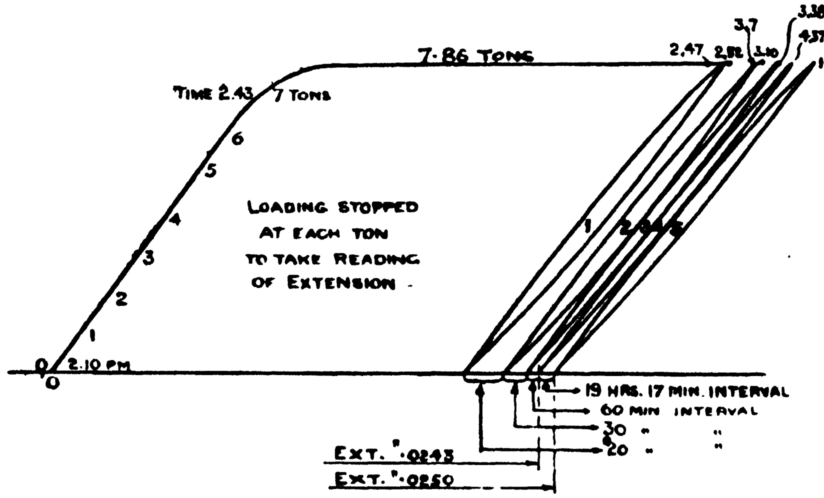
Each small circle denotes a loop, and the letter written against some of them identifies the first loop on the plate corresponding with the letter. The subscript 1 means the 1st loop.

To face p. 132.

Processions of loops were similarly taken from a nickel steel test piece, from a mild steel test piece, and from an iron test piece.

### § 9. *Looping under Constant Load.*

The diagram in fig. 24 shows the effect of looping under constant load.



| LOOP | AREAS | |
|---|---|---|
| 1 | 0·25 sq. in. | PL. 562 A |
| 2 | 0·18 " " | |
| 3 | 0·155 " " | |
| 4 | 0·156 " " | |
| 5 | 0·17 " " | PL. 562 B |

— LOOPING UNDER CONSTANT LOAD FIG. 24

— 0·8 CARBON STEEL —

The test piece was placed in a Buckton testing machine and loaded gradually until yield began at 7·86 tons, and then loading was stopped. The extension was allowed to proceed under this load for 4 minutes. Then the load was removed and re-applied and the spot of light traced out loop 1.

Extension continued slowly under the load, still maintained at 7·86 tons, and loops 2, 3 and 4 were taken at intervals of 20, 30 and 60 minutes respectively.

The next interval was 17 hours 20 minutes, during which time the gauge length extended $\frac{1}{1000}$ inch approximately. Loop 5 was then taken. Comparing these loops it will be seen that there is no recovery of proportional elasticity although the piece was allowed to stretch under the yield load of 7·86 tons for about 20 hours.

This shows that if the yield load is kept on until the test piece has stopped extending, a process which may take a long time, at the end of the experiment the test piece will not have gained proportional elasticity. It is still in a state of imperfect, or unproportional elasticity.

### § 10. *The Practical Utility of the Load Elastic Extension Looped Diagram.*

A diagram showing the elastic line and a few loops is of great practical utility in industrial applications. The data immediately measurable from the diagram are :—

(1) the load at the limit of proportionality;

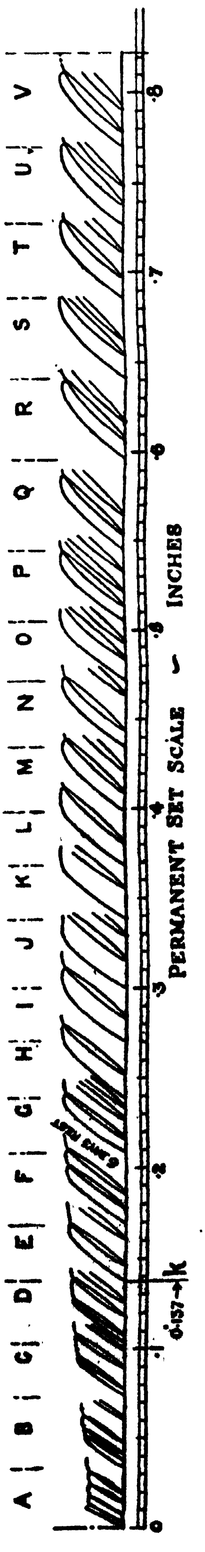


Fig. 19

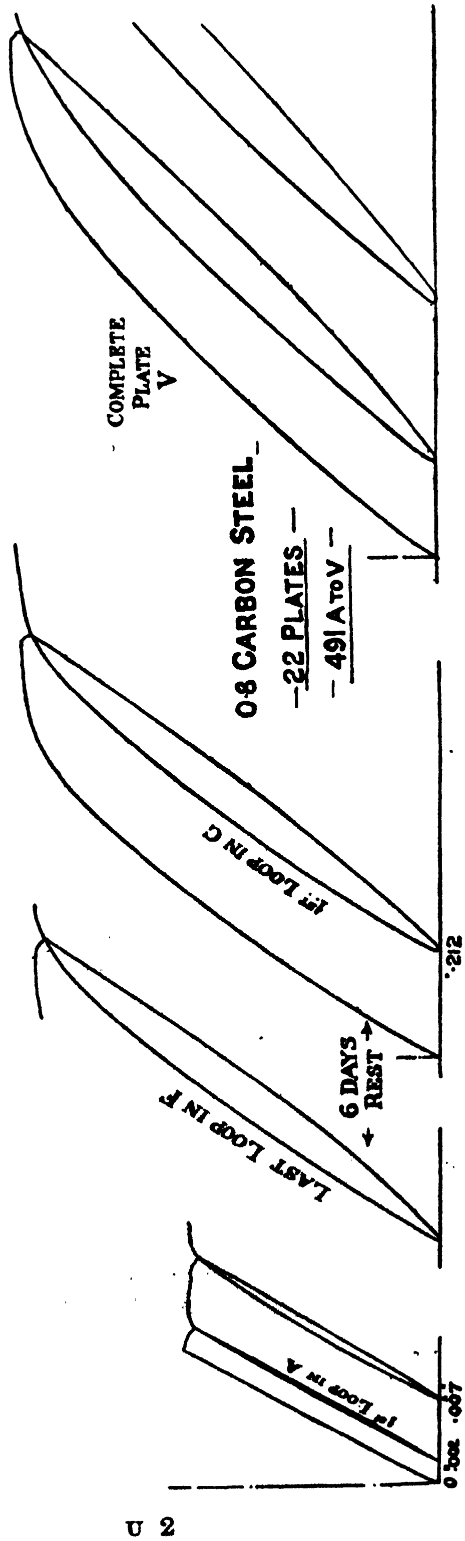


Fig. 20

Fig. 21

Fig. 22

Fig. 23

(2) the yield load;
(3) the value of E: this is given by the slope of the elastic line;
(4) the work lost per loop;
(5) the rate of increase of the work lost per loop.

The diagram from *normal material* satisfying known conditions of composition and manufacture may be used as a diagram of comparison. The form of the curve, the loop area and the rate of increase are sensitive to changes in the kind of material and to changes in the inner state of materials.

The diagram is specially useful in showing the load at the limit of proportionality, for this load bears neither a constant relation to the yield point, when there is one, nor to the ultimate load. Consequently factors of safety reckoned against either the yield load or the ultimate load are ambiguous.

This is specially important in gun design. The whole theory rests upon the elastic property of the material, and the theory ceases to apply after the limit of proportionality is passed.

Considerable research in many directions is necessary before a full interpretation can be given to the looped diagrams, and for the present I will reserve further discussion.

### § 11. *Correlation of Diverse Tests by the Load Extension Diagram.*

Load extension diagrams of the kind shown in this and former papers are likely to be useful to the engineer and metallurgist in the correlation of the many different tests now made to ascertain the quality of metals.

For example, fig. 25 was taken from a test piece of material giving a low impact number. Its shape differs markedly from the shape of a normal diagram. It corresponds in fact with the shape of the curve found from overstrained material.

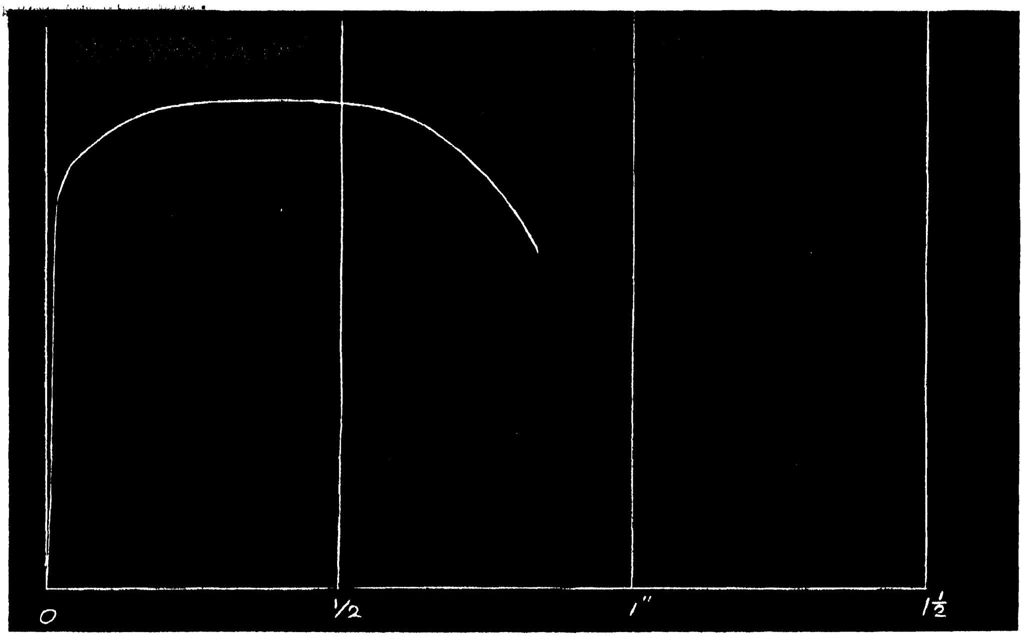


Fig. 25 (overstrained mild steel).

The inference is that the material is in the overstrained condition. That this inference is correct is shown by fig. 26. A test bar of the same material was annealed by heating to 550° C., and then cooling slowly within the furnace, and then it was found that it gave the diagram of fig. 26, which is the normal shape for the class of steel tested.

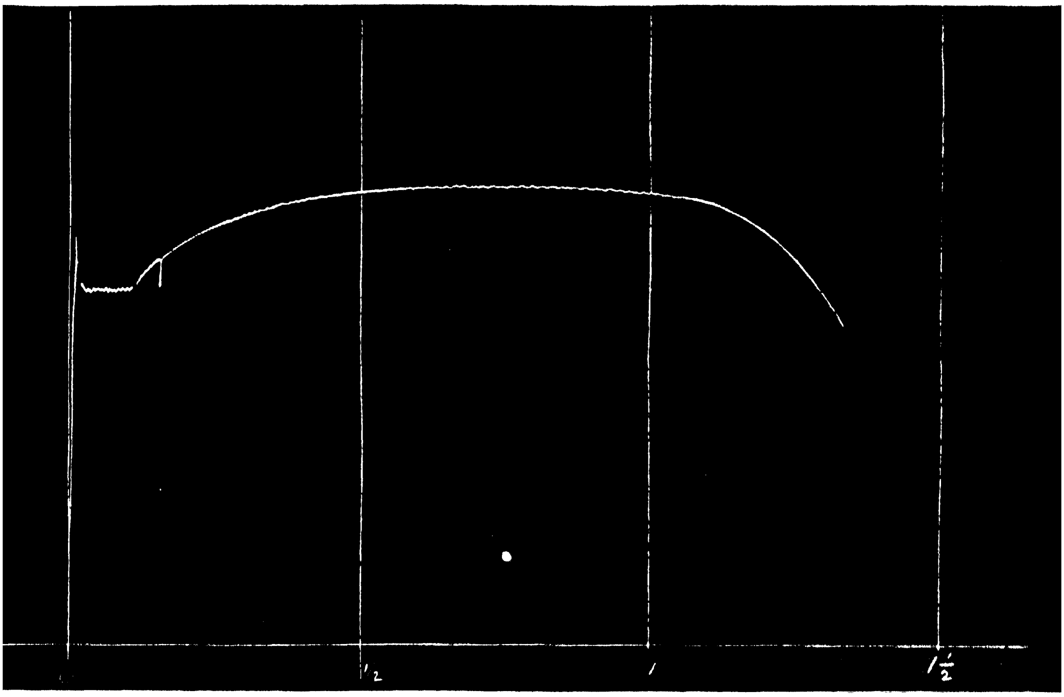


Fig. 26 (normal mild steel).

I have tested many bars of steel rejected on shock test and giving low impact numbers, and I have always found that the shape of the load extension diagram discloses the abnormal state of the metal. Much work of a comparative kind must be done before the result can be widely generalized. The War Committee of the Royal Society did a considerable amount of work in this direction.

The limiting fatigue stress may probably be found by inspection from a load elastic extension diagram.

It is probable that *the limiting range of stress in fatigue has for its positive value the stress equal to the limit of proportionality.*

Referring to the diagram for iron, fig. 7, it will be seen that the limit of proportionality is at about 3½ tons, corresponding to 11·45 tons per sq. inch.

I prepared six test pieces of the material, and Dr. Stanton kindly applied his fatigue test to them at the National Physical Laboratory. He found, after applying alternating loads in the aggregate 24,000,000 times to the eight test pieces, that the approximate limiting range of stress in fatigue was between ±10½ and ±13 tons per sq. inch. The average is ±11·75 tons per sq. inch.

The agreement between the limit of proportionality shown on the diagram, namely, 11·45 tons per sq. inch, and the fatigue limit found by quite a different test, is remarkably close.

Again, if this result could be generalized, it could be asserted that every load elastic extension diagram shows the positive value of the fatigue limit. The long and tedious experiments with alternating loads would be unnecessary. Such a conclusion requires comparison to be made over a wide range of material. This, therefore, is a promising field of research.

These brief notes of results and inferences show what a wide range of information lies before the engineer and metallurgist if he has before him a pair of diagrams, the one showing the complete load extension curve from zero to fracture, the other a load elastic extension looped diagram on a large extension scale.

The matter incorporated in this paper has been selected from experiments extending over several years. I desire to express my acknowledgments and thanks to Prof. Witchell for his help, and in particular for the assistance he gave me in reducing the looped diagram to the curves of Sheet 1.

I also desire to acknowledge the assistance of Mr. Orr for the skill and care with which he has drawn the curves on Sheet 1 and figures 19–23 from the photographic plates.

# V. *The Stress-strain Properties of Nitro-cellulose and the Law of its Optical Behaviour.*

*By Prof.* E. G. COKER, *F.R.S., and* K. C. CHAKKO, *M.Sc., University College, London.*

Received September 22, 1919,—Read January 22, 1920.

THE physical characteristics of transparent bodies capable of resisting stress have been the subject of much investigation, and in particular the properties of various glasses* have been studied with much thoroughness since these latter have an extensive use both for commercial and scientific purposes.

In recent years many new forms of optical materials have found an industrial use, and especially nitro-cellulose compounds, which are valuable in cases where glass is not suitable.

The mechanical and optical properties of such bodies have not, so far, been examined in very great detail, and the present paper describes some experimental evidence which has been obtained and which it is hoped to extend as opportunity occurs, since this has an important bearing on the study of stress problems arising in engineering practice.

The principal matters which are examined in the present communication are the mechanical properties of nitro-cellulose under pure tensile and bending stresses and the laws of its optical behaviour under these kinds of stress. In the course of the experimental study a considerable number of specimens have been examined, all of which are the manufacture of the British Xylonite Company.

The salient features of the material are its great flexibility and toughness, and the ease with which it can be drilled, turned or machined. By suitable adjustment of the condition of nitration of the body the hardness of the material can be varied through a considerable range, but owing to the difficulties created by the stress of war it has not been possible to make this investigation cover materials possessing a great range of hardness, and in fact all the specimens taken were originally selected

---

* L. N. G. FILON, "On the Variation with the Wave-length of the Double Refraction in Strained Glass," 'Camb. Phil. Soc. Proc.,' vol. XI., Part VI.; vol. XII., Part I.; and vol. XII., Part V.; see also 'Phil. Trans.,' A, vol. 207, and 'Roy. Soc. Proc.,' A, vols. 79 and 89. F. POCKELS, "Über die Aenderung des Optischen Verhaltens verschiedener Gläser durch elastische Deformation," 'Annal. d. Physik,' 1902, and F. D. ADAMS and E. G. COKER, "The Cubical Compressibility of Rocks," 'Trans. Carnegie Institute.'

[Published October 18, 1920.

on account of their transparency and freedom from initial stress. The sheets from which the specimens are made differ greatly in age, one has been in stock for at least eight years, most of the others have been stored one or more years.

In order to examine the stress-strain properties of this material it is unnecessary to use a very delicate extensometer as the value of the modulus for direct stress is comparatively small, and for the purposes of these experiments a very simple form is employed consisting of a pair of clips attached respectively to a scale and a pointer, which latter slides over the scale and is kept in contact with it by suitable attachments. In order to examine the optical properties of the material while under stress, both scale and pointer are perforated to give a window opening, and thereby permit a beam of polarised light to be transmitted through the specimen under examination. With this instrument and special magnifying devices it is possible to estimate extensions of 0·0002 inch.

A preliminary examination of the problem set out above may be described with reference to some experiments on a specimen which was originally used in 1911 for determining the stresses in a notched tension member.*

Two bars, each 1 inch wide, were cut, at that time, from a clear plate of xylonite $\frac{3}{16}$ inch thick, and each was fashioned with notches of different sizes along the edges. One of these specimens has been used for the present test. A length of 6 inches was used for observations of the longitudinal strains, while the lateral strains required for determining POISSON'S ratio $\frac{1}{m} = \sigma$ have been measured by aid of a strain-measuring apparatus having a unit reading of $\frac{1}{1,000,000}$ inches. Unless otherwise stated these latter measurements are in the direction of the thickness of the material. As the details of the measuring apparatus and the cylindrical recorder used with it have already been described† they are not referred to further here.

*Longitudinal Extension.*—The specimen was examined in the polariscope under a moderate load and it was found that the stress was very uniformly distributed over a length of $6\frac{1}{2}$ inches, but the remaining part of the parallel portion showed signs of unequal stress distribution owing to the enlarged ends. It was therefore marked off approximately into half-inch lengths over a total length of 6 inches, the exact distance being read to $\frac{1}{1000}$ inches.

YOUNG'S *Modulus.*—As it is convenient to start with a load of 20 lbs. on the specimen, a preliminary observation is made to determine the corresponding extension, and this value is allowed for in subsequent readings for convenience in plotting from a zero strain value.

---

* E. G. COKER, "The Effects of Holes and Semi-circular Notches on the Distribution of Stress in Tension Members," 'Proc. Phys. Soc.,' 1911.

† "Photo Elasticity for Engineers," by Prof. E. G. COKER, D.Sc., F.R.S., 'The Institute of Automobile Engineers,' November, 1917.

Six sets of observations are shown in fig. 1, the loads being applied in 20-lb. increments.

The maximum loads vary from 160 lbs. to 200 lbs., and after the strains reach their final value the load is again taken to its initial value and the scale read, to note the "semi-permanent set." The time interval between successive loads varies from one to two minutes. It is found that the stress-strain readings so obtained are approximately linear, except at the highest values, but in order to obtain as correct a value of YOUNG'S Modulus, E, as possible only measurements between 40 and 140 lbs. load are used, as the readings between these points are considered to be the most reliable. The strain corresponding to this difference of stress of 1256 lbs. per sq. inch is 0·00354 inch giving a value of E = 355,000.

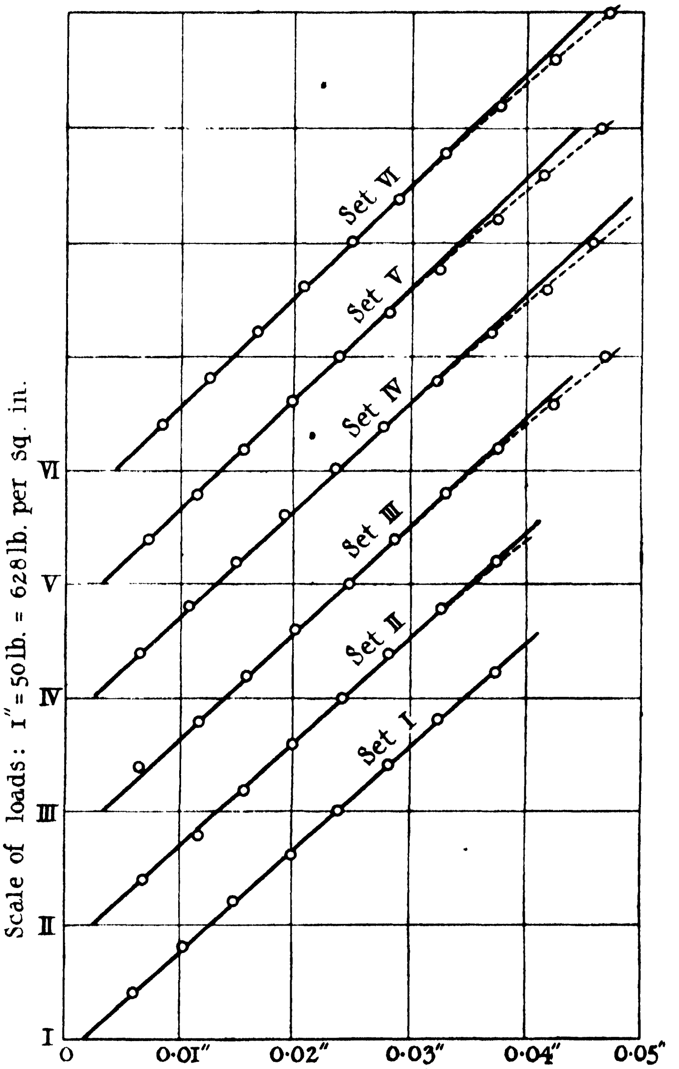


Fig. 1.

It will be observed that there is no very pronounced elastic limit, and that the curve is nearly straight up to 150 lbs. load (1900 lb./in.$^2$), which latter value may be taken as the elastic limit of the material. There is a "semi-permanent" set of 0·001 inches for each repetition of load, and a pronounced recovery between successive loadings especially with a short period of rest.

*Measurements at Higher Stresses.*—The spring balance used to measure the moderate loads in the above observations had a maximum capacity of 200 lbs., but for the higher stresses required, a balance recording up to 500 lbs. was necessary, the observations being made in a similar manner with readings on the magnified scale up to 400 lbs. load, and after this coarser readings were taken with the telescope. A maximum load of 476 lbs. (6000 lb./in.$^2$) was reached, but as the extension then increased very rapidly it was not possible to keep the load at this maximum value, moreover, as the stressing frame was of rather limited capacity for large strains, the test could not be carried to fracture, although a total extension of 1·211 inches was obtained. The observations also showed that the permanent extension was very uniform from section to section.

The condition of the material has in fact some resemblance to that of a mild steel which has been overstrained and allowed to rest. This is shown by a subsequent experiment in which the loading was repeated and the stress-strain properties examined anew. It was then found that the elastic limit of the material was still approximately at 150 lbs. load, corresponding to a stress of 1900 lbs. per sq. inch, but the modulus E had now risen to 502,000, measured in pound and inch units, as a result of the overstrain. The relations of load to extension for both conditions are shown in the accompanying fig. 2, but as the scales of load and extension are the same for both experiments the curve for overstrained material lies below that for unstrained material. It may be observed that the material possesses, in a marked degree, the property of contraction when the load is removed even when very much overstrained, and in this case when the full load of 300 lbs. was removed the semi-permanent extension was only 0·006 inch, and half of this disappeared with a few minutes rest.

Observations of lateral strain were also made with a suitable extensometer at several sections of the test bar, and their mean value for 100 lbs. load showed a strain of 0·00144, corresponding to a value of $m = \frac{1}{\sigma} = 2·45$ where $\sigma$ is POISSON'S ratio.

The value of E is high as later experiments show, and this may possibly be due to an ageing effect, as in process of time the material appears to undergo some change, especially if the cut surfaces are not highly polished. This may probably be ascribed to the escape of a small portion of the volatile constituent of the material. It is also worthy of remark that the usual method of polishing appears to produce a thin outer layer which is harder than the interior, and this also has the effect of raising the value of E in thin specimens.

The effect of removing this thin layer of hard material has been under observation for some time, but in the experiments described here the flat sides are untouched, and the cut edges are unpolished although quite smooth.

*Optical Properties.*—There is considerable colour when an over-strained specimen is examined under no load in the polariscope. In the parallel part the colour is very uniform, and by comparing this with a previously unstrained specimen under load it appears that the permanent colour indicates a complex state of stress, since it could not be completely neutralised by a comparison tension piece, nor by bending the strained bar itself.

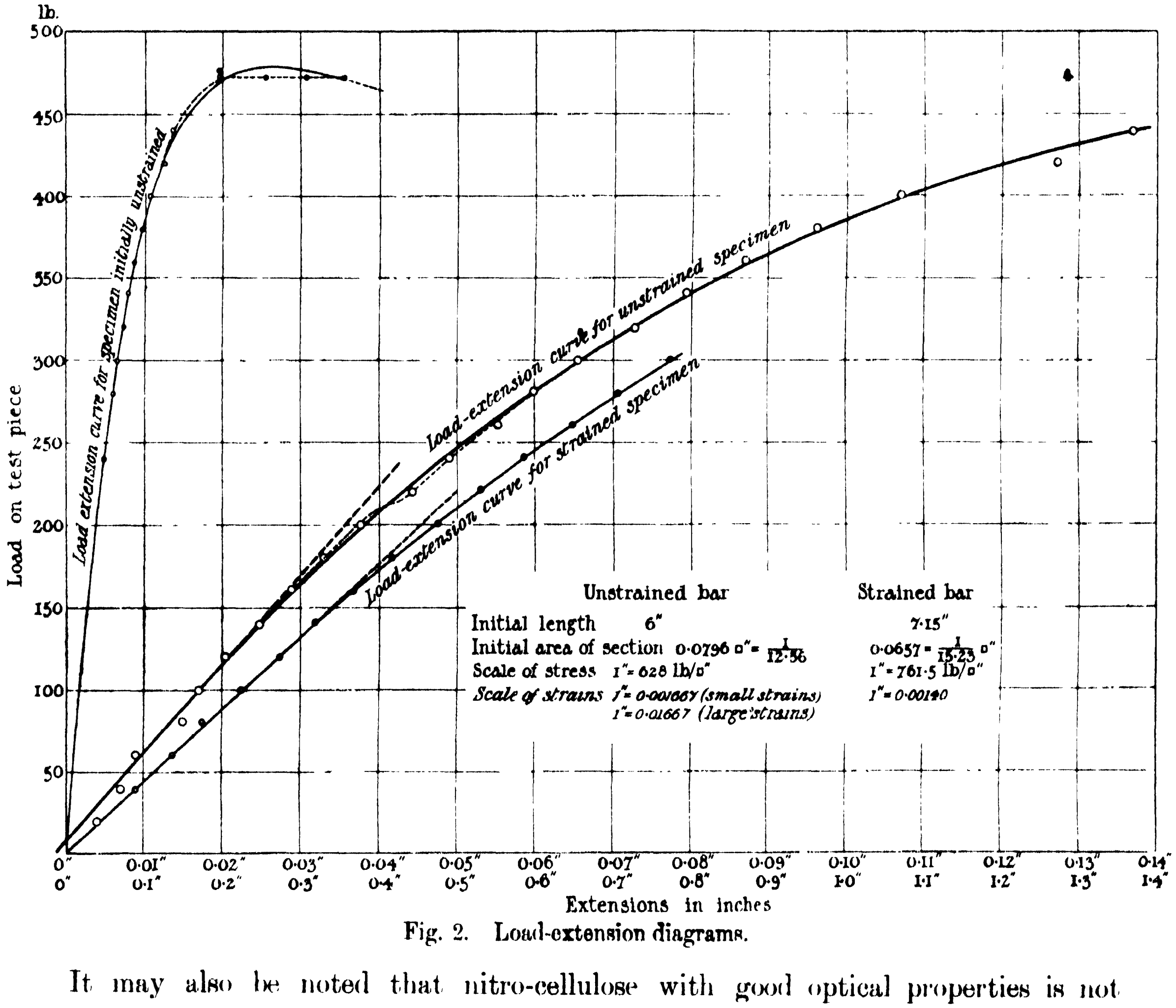


Fig. 2. Load-extension diagrams.

It may also be noted that nitro-cellulose with good optical properties is not apparently procurable, above ¼ inch thick, and it is difficult therefore to conform to the laws of similarity for the test specimens used in this investigation.

In later experiments the stress-strain properties of the material are examined both below and above the elastic limit, and the values of YOUNG'S modulus and POISSON'S ratio are measured for a number of specimens of different thickness and varying age. Especial attention is also directed to test the validity of the stress-optical law of this material since this is a matter of fundamental importance and little attention has so far been given to it.

The result of breaking the first three specimens showed that the extensometer arrangement was defective. The tiny indentations at the sides, where the extensometer was attached, so weakened the specimens that they all broke at one or other of these sections. The thinnest one naturally was affected most, so much so that it fractured with very little extension. This defect was partly corrected by cementing small fillets on to the specimen from which the extensometer clips were supported, but in spite of this some of the specimens fractured outside the gauge limits, showing that even in a ductile material the effects of enlarged ends prevents equalisation of stress near the change of section under any condition of load.

*Stress Optical Determination.*—It was originally intended to study the stress-optical properties of nitro-cellulose by analysing the light which traversed the material by means of a spectroscope ; but the necessary apparatus was rather difficult to procure, and it was convenient therefore to commence with a standard nitro-cellulose beam and use this for comparison with the optical phenomena observed in tension. The methods adopted here proved to be exceedingly well adapted for measurement of stress distribution beyond the elastic limit and are likely to be of great use hereafter. The comparison beam used is of rectangular section and is subjected to pure bending moment of known amount, and the stress at any point can therefore be calculated from the formula $f = \frac{My}{I}$ without appreciable error.

It is generally assumed that the relative retardation of the polarised rays in a piece of optical material under moderate stress is proportional to the difference of principal stresses at the point, but this may not be correct and cannot be assumed to hold without experimental proof. Hence the stresses in the comparison beam are restricted to small values, so that the limit of proportionality of stress to strain is not passed in order to give an opportunity of examining the possibility of the law following a linear strain function or possibly some more complex variable. In order to make the retardation in such a beam sufficiently great to balance the retardation in the highly stressed specimens, the thickness of the beam should be large. This is most conveniently obtained by placing the several beams side by side, with their ends clamped and pinned together, as shown in fig. 3 in which several beams are so fastened together by plates A, to which extension levers B are also attached for supporting loads C depending from hangers D. This compound beam is supported on knife edges two inches apart, and when loaded has its central section sufficiently removed from the supports to give pure bending moment at the central section. The material of the beams is almost perfectly elastic up to and probably beyond 1600 lb./in.$^2$, but they are actually not stressed to more than 1300 lb./in.$^2$. In some cases as many as eight beams $\frac{1}{8}$ inch thick are used in this way, and a strong beam of light is then necessary to enable a comparison to be made with the tension member under observation. A carbon arc is then used as the source of light, but when only two or three thicknesses are employed the light from a Nernst lamp is sufficient, but in all cases the images are observed directly by eye instead of projecting on to a screen. The general arrangement of the apparatus is shown in the accompanying fig. 4, in which a plane polarised beam of white light from a NICOL'S prism A is transmitted through the tension specimen B, to which an extensometer C is secured, and is then focussed by a lens D on a horizontal slit in order that the light passing through the comparison beams shall be at the same level throughout. This thin pencil of light is again brought to parallelism before passing through the compound beam F and analyser G, and is finally focussed on a ruled glass slide H provided with an eye-piece J. The weight of the extension beams and hangers causes a bending moment in the beams which has been allowed for in all calculations of stress. In order to compare the different specimens one with another, an "equivalent stress" in each specimen is calculated, that is, such a stress as would produce the same relative retardation in a piece of nitro-cellulose of the same material as the standard beam, but of the thickness of the specimen under observation. Thus if the thickness of specimen is $t$ and the stress in

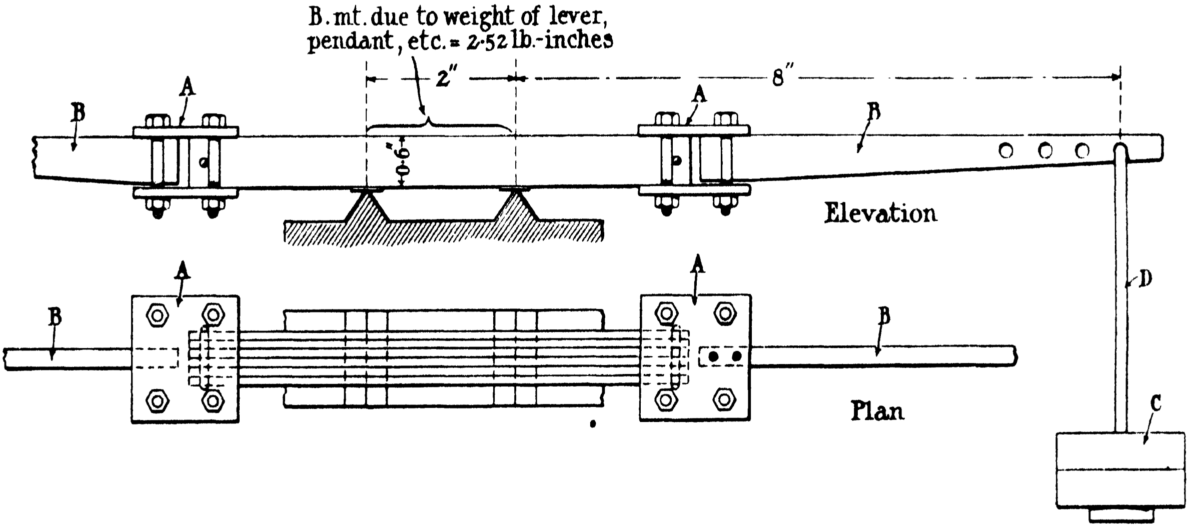

B.mt.due to weight of lever, pendant, etc. = 2·52 lb.-inches
2″
8″
A
B
A
B
0·6″
Elevation
A
B
A
B
D
Plan
C

Fig. 3. Beam Comparator.

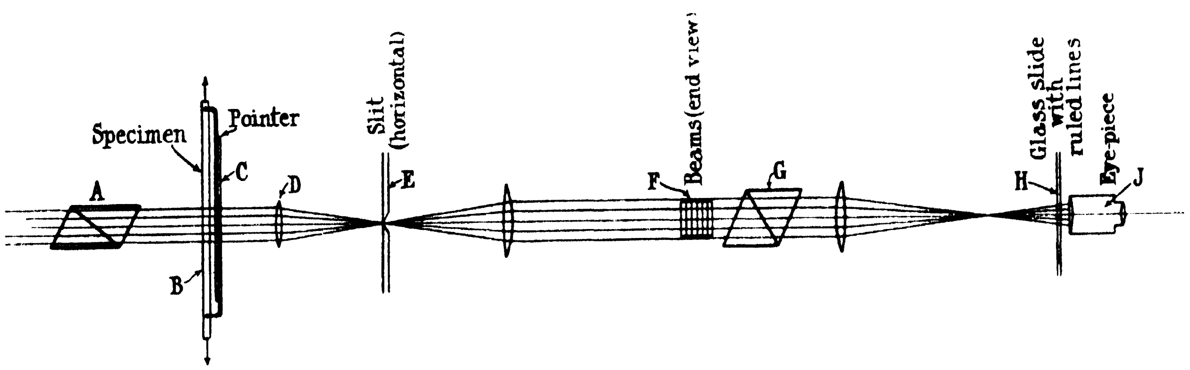


Fig. 4. Diagram of polariscope.

the beam at the points where the colour in the specimen is neutralised is $f_0$ and the corresponding thickness is $t_0$, we have the equivalent stress in the specimen $\frac{f_0 t_0}{t}$.

Now if M is the bending moment in the beam, $d$ is its depth, and $y$ is the distance from the neutral axis, then

$$f_0 = \frac{My}{I} = \frac{My}{\frac{1}{12}t_0 d^3} = \frac{12My}{t_0 d^3},$$

so that the equivalent stress

$$f = \frac{f_0 t_0}{t} = \frac{12My}{td^3}.$$

Although as stated above the law of optical retardation is generally assumed to follow a linear law of stress difference, yet there is no apparent reason why it should not follow some other law, as for example a linear strain law or possibly contain terms involving squares of stress or strain. Some attempt has been made to find if the latter assumptions have any foundation, but if so the effects are within the limit of experimental error, and too small to be of any significance with the effect produced by a linear relation.

As regards the question whether this relation should be expressed in terms of stress or strain, it may be pointed out that an attempt is made here to test this with materials under direct stress, and that the validity of the law for combined stresses and strains still remains for consideration (apart from lateral strains, which are presumed to have no effect beyond altering the length of the path in which retardation takes place), but as in this case, if the standard, not stressed beyond the elastic limit is compared with another in which this condition is passed the experiments do in fact provide a means of discrimination, since in the standard, stress and strain are proportional, but are not so in general for the tension member. Hence if the form of the law of optical effect is assumed in terms of stress it does not exclude the possibility of finding from the experimental evidence whether it should not be expressed in terms of strain. We may, therefore, without loss of generality, take as an assumption the usual relation that relative retardation $R = C(P-Q)t$ as a convenient expression where $(P-Q)$ is the difference of principal stress $= f$, $t$ is the thickness of the material and C is the stress-optical coefficient.

Let $C_0$ be the stress-optical coefficient of the standard beam. Then $R_0 = C_0 f_0 t_0$ for this beam.

When this latter is used to neutralise the retardation R in the specimen, since $R = R_0$, we have

$$Cft = C_0 f_0 t_0$$

But $t = t_0$, here and therefore

$$Cf = C_0 f_0.$$

Now in general there is an initial retardation which is independent of any load. Let this correspond to stresses F, $F_0$. Then the condition $Cf = C_0f_0$ becomes

$$C(f+F) = C_0(f_0+F_0)$$

or differentiating,

$$C \,.\, df = C_0 \,.\, df_0,$$

therefore

$$\frac{C}{C_0} = \frac{df_0}{df} = \frac{1}{\dfrac{df}{df_0}} = \text{reciprocal of slope of the stress/equivalent stress,}$$

which affords a convenient relation for examining the experimental data.

Turning now to the further experimental data upon the stress-strain properties of nitro-cellulose in tension a number of experiments have been made upon material of varying age and thickness, and these are plotted in the accompanying fig. 5, to show their characteristic properties under loads which sometimes exceed very considerably the elastic limits of the material.

With the thinnest specimens $\frac{1}{16}$ inch thick it was not found possible to obtain a reliable value of POISSON'S ratio, but YOUNG'S modulus, E, has been found, and the measurements plotted in fig. 5 show a characteristic feature that, although the first test is carried well beyond the elastic range, as soon as the load is removed a total extension of 0·0608 inch is reduced to 0·0070 inch or only 0·001 inch more than obtained at the commencement of this test. Moreover the value of the modulus changes less than 2 per cent. under these circumstances due to the earlier loading. It is also large as the skin effect is pronounced. These general characteristics are also observable in the measurements recorded in this figure for much thicker material if due allowance is made for the diminished effect of the surface layers. The capacity of returning to its original shape after high loads is still more marked in the next series of experiments on material $\frac{1}{8}$ inch thick, fig. 6, in which a stress of nearly 5000 lbs. per sq. inch is reached in the first experiment (curve 1) with nearly complete recovery, and when further loads with maxima varying from 4000 to 5000 lbs. per sq. inch are applied (curves 2 to 8) these give almost identical values of YOUNG'S modulus on the straight part of the curve until the ninth loading, where there is a sudden fall to E = 261,000 with an extension of 0·1045 inch corresponding to the initial load of 20 lbs. After this, with the considerable initial extension of 0·4290 inch, there is a great rise in the modulus. The value of POISSON'S ratio is very constant and is here found to reach the highest value of $m = \frac{1}{\sigma} = 2{\cdot}7$.

Succeeding experiments on still thicker material confirm these results, and with the exception of the $\frac{3}{8}$-inch plate, the load extension curves agree in their linear character up to about 2000 lbs. per sq. inch, although the specimens differ in age and possibly also in composition. They have, however, the common feature of possessing

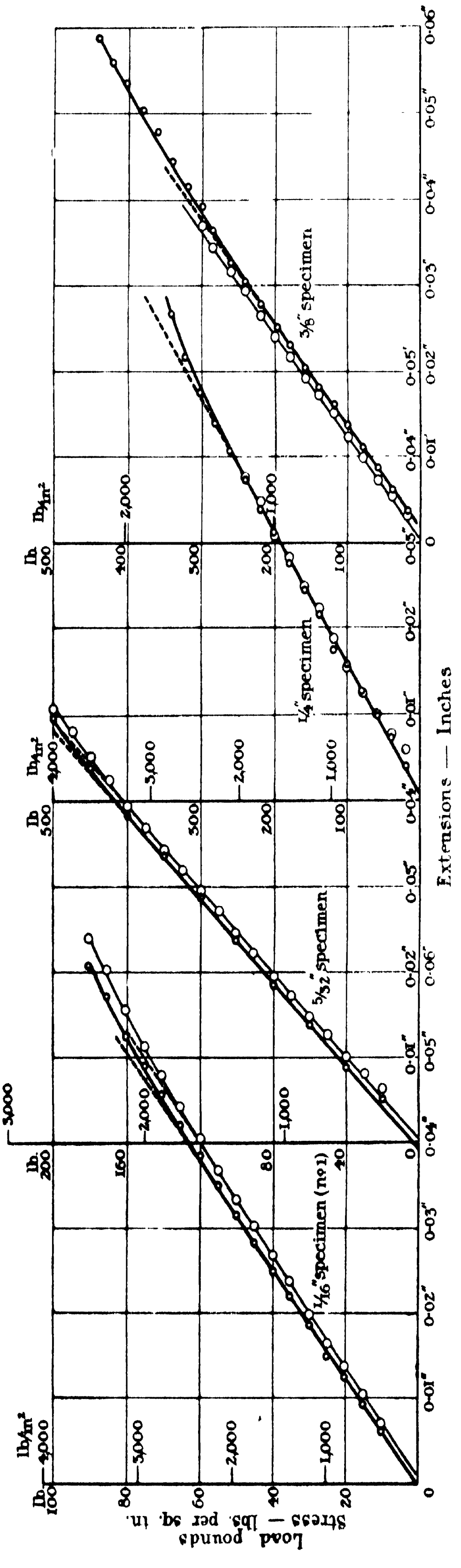


Fig. 5.

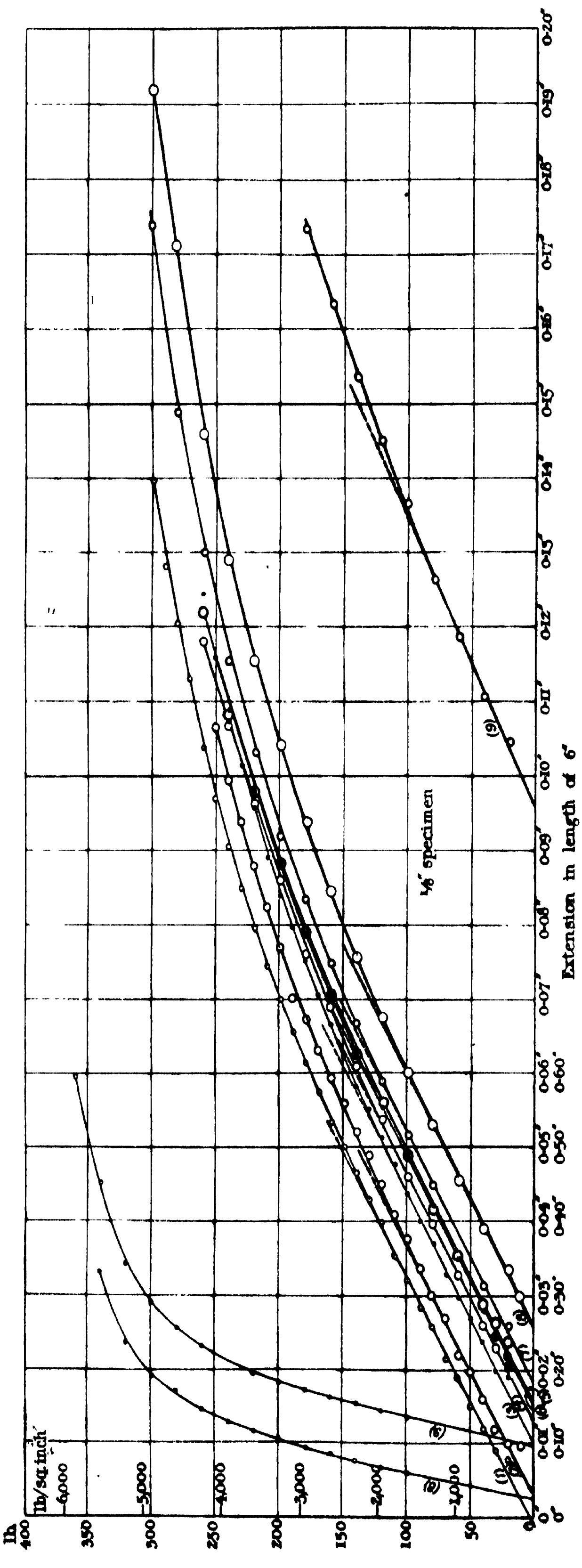


Fig 6. Load-extension diagram—low stresses.

excellent optical properties and freedom from initial stress. The thickest plate, however, is exceptional, as its optical properties are poor.

Further experimental work on these materials is almost entirely devoted to the examination of the optical law of retardation under load, and for convenience all the data which follows is expressed as a stress or a strain, the units being pounds and inches, in which $e$ is the strain under direct stress $f$, and the equivalent stress $f_0$ is obtained from the comparator beam. A typical example of these values is given in Table I., for material ¼ inch thick, as these measurements are referred to later for comparison with values of stress and strain obtained from spectrum observations.

TABLE I.

| ¼-inch. | | | ¼-inch. | | |
|---|---|---|---|---|---|
| $e$. | $f$. | $f_0$. | $e$. | $f$. | $f_0$. |
| 0 | 0 | 73 | 0·0123 | 3315 | 3440 |
| 0·0003 | 160 | 220 | 0·0135 | 3505 | 3560 |
| 0·0007 | 319 | 322 | 0·0142 | 3665 | 3900 |
| 0·0013 | 478 | 440 | 0·0153 | 3825 | 4140 |
| 0·0020 | 637 | 660 | 0·0172 | 3980 | 4500 |
| 0·0027 | 797 | 880 | 0·0187 | 4140 | 4770 |
| 0·0033 | 956 | 1025 | 0·0208 | 4300 | 5100 |
| 0·0037 | 1115 | 1173 | 0·0212 | 4460 | |
| 0·0043 | 1274 | 1320 | 0·0237 | 4620 | |
| 0·0050 | 1433 | 1495 | 0·0287 | 4780 | |
| 0·0055 | 1594 | 1642 | 0·0320 | 4940 | |
| 0·0058 | 1752 | 1760 | 0·0370 | 5100 | |
| 0·0063 | 1912 | 1910 | 0·0553 | 5260 | |
| 0·0070 | 2070 | 2050 | 0·120 | 5420 | |
| 0·0077 | 2230 | 2200 | 0·157 | 5580 | |
| 0·0082 | 2390 | 2350 | 0·183 | 5740 | |
| 0·0087 | 2550 | 2540 | 0·247 | 5900 | |
| 0·0095 | 2710 | 2310 | 0·280 | 6055 | |
| 0·0102 | 2865 | 2860 | 0·340 | 6215 | |
| 0·0108 | 3025 | 3080 | 0·357 | 6375 | |
| 0·0117 | 3185 | 3230 | | | |

In the earlier experiment on the material $\frac{1}{16}$ inch thick, a fracture was obtained near the change of section and before the full extension developed, but still very nearly at the full load. It is included here (fig. 7) as, although the later parts of the stress-strain curves are not entirely satisfactory, this does not affect the problem in hand, since the stress-strain curve is not required very much beyond a pronounced yield in the material.

As a purely mechanical problem, however, there is a considerable amount of interest attaching to the accurate measurement of stress and strain over the whole of the plastic region, and it may be worth while at some future time to examine this with some care, especially if optical methods are applied to study the distribution of stress in purely plastic materials.

The stress-strain curves obtained for this thin material show a divergence from a linear law above 2000 lbs. per sq. inch whether plotted from the direct load or the optical stress measurements, but if the direct stress is plotted against its optical equivalent there is a definite linear law extending up to at least 4500 lbs. per sq. inch, and with only a small divergence at 5000 lbs. per sq. inch. The results in fact go to show that the law of retardation is linear as regards stress not only up to the elastic limit but actually to at least twice this range, where it is quite impossible for the strain to be linear. This result is shown in all the experiments on good optical material. Thus in plates $\frac{1}{8}$ inch thick where the elastic limit appears to be about 2250 lbs. per sq. inch, fig. 8, the corresponding value for $f/f_0$ shows no divergence from linearity until nearly double this amount, although the curve of $f_0/e$ ceases to be linear at about 2500 pounds per sq. inch.

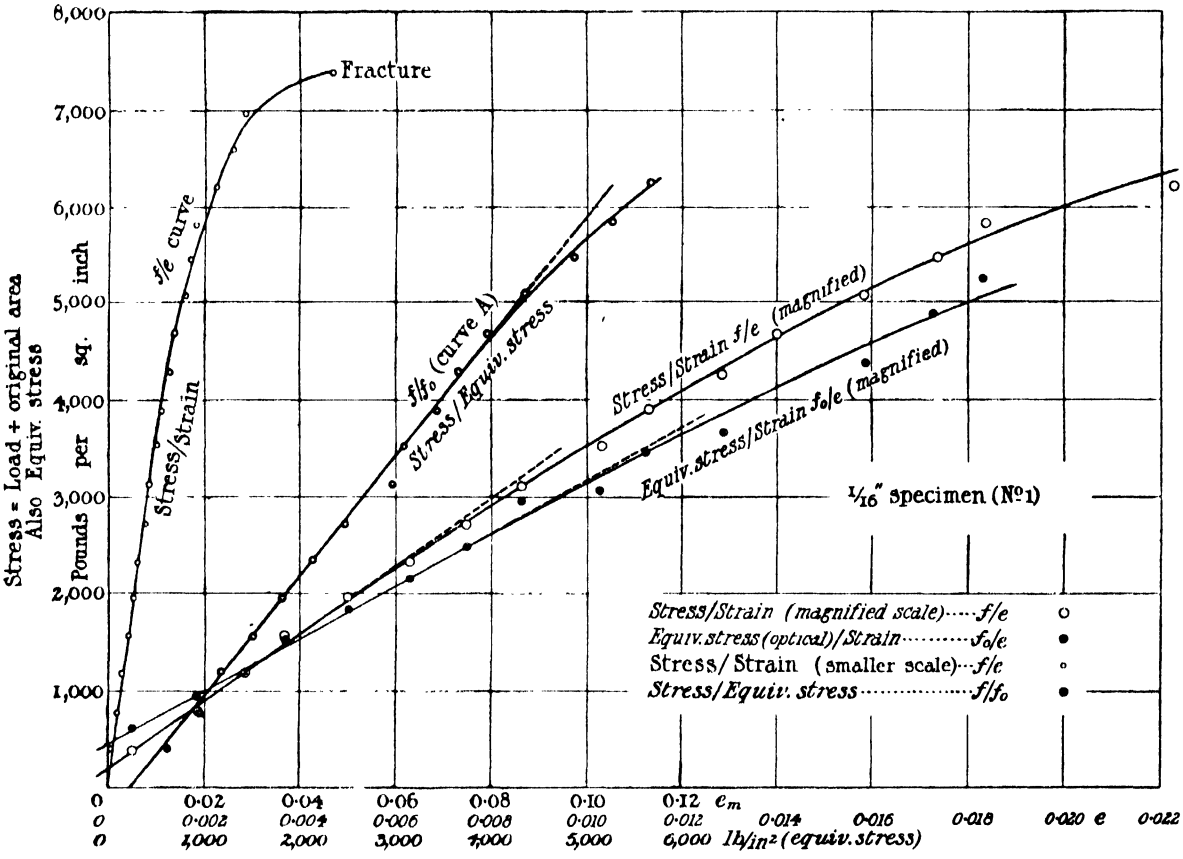


Fig. 7. Stress-strain curves of nitro-cellulose.

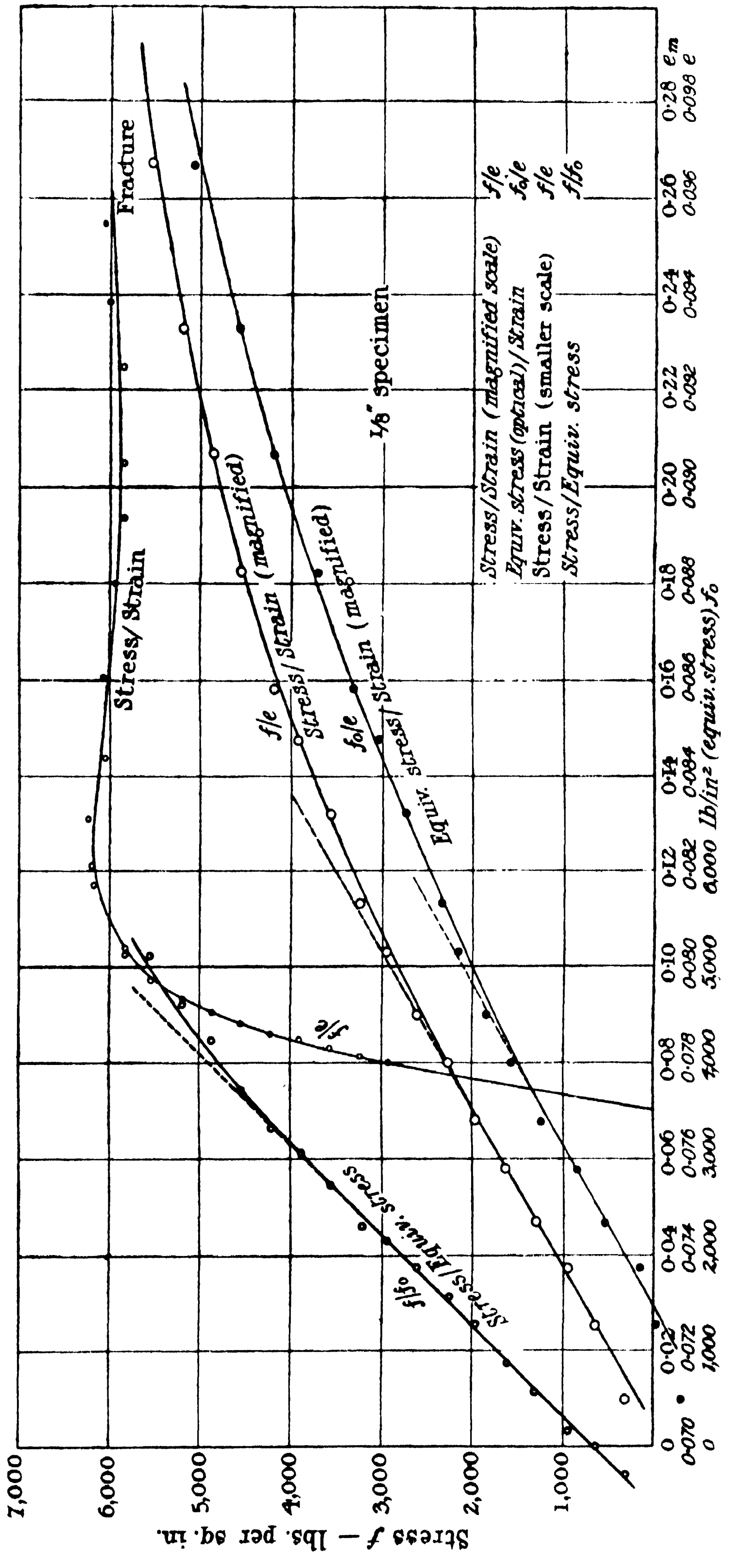

Stress $f$ — lbs. per sq. in.
0
1,000
2,000
3,000
4,000
5,000
6,000
7,000
0 0·02 0·04 0·06 0·08 0·10 0·12 0·14 0·16 0·18 0·20 0·22 0·24 0·26 0·28 $e_m$
0·070 0·072 0·074 0·076 0·078 0·080 0·082 0·084 0·086 0·088 0·090 0·092 0·094 0·096 0·098 $e$
0 1,000 2,000 3,000 4,000 5,000 6,000 lb/in² (equiv. stress) $f_0$
Fracture
Stress/Strain
$f/e$ Stress/Strain (magnified)
$f_0/e$ Equiv. stress/Strain (magnified)
1/8" specimen
$f/f_0$ Stress/Equiv. stress
$f/f_0$
Stress/Strain (magnified scale) $f/e$
Equiv. stress (optical)/Strain $f_0/e$
Stress/Strain (smaller scale) $f/e$
Stress/Equiv. stress $f/f_0$

Fig. 8.

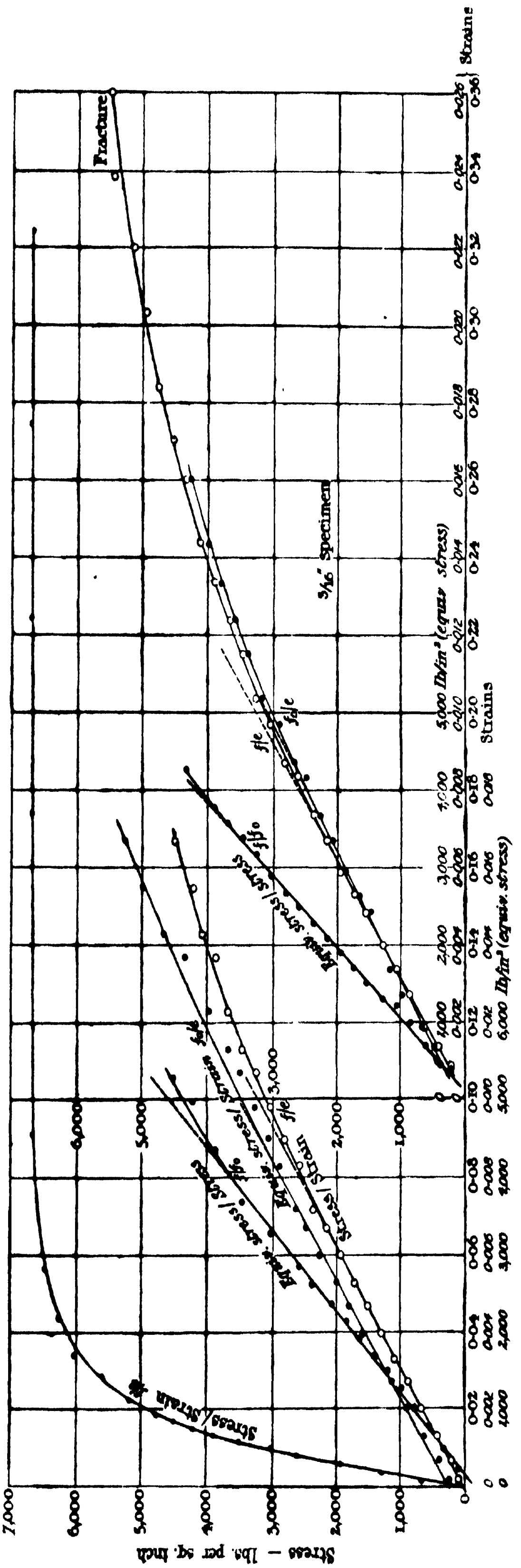

Stress – lbs. per sq. inch
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
Strains
0
0·02
0·04
0·06
0·08
0·10
0·12
0·14
0·16
0·18
0·20
0·22
0·24
0·26
0·28
0·30
0·32
0·34
0·36
lb/in² (equiv. stress)
Fracture
3/16" Specimen
Stress/Strain f/e
f/e
f₀/e
Basic stress/strain
f/f₀
Stress/stress

Fig. 9.

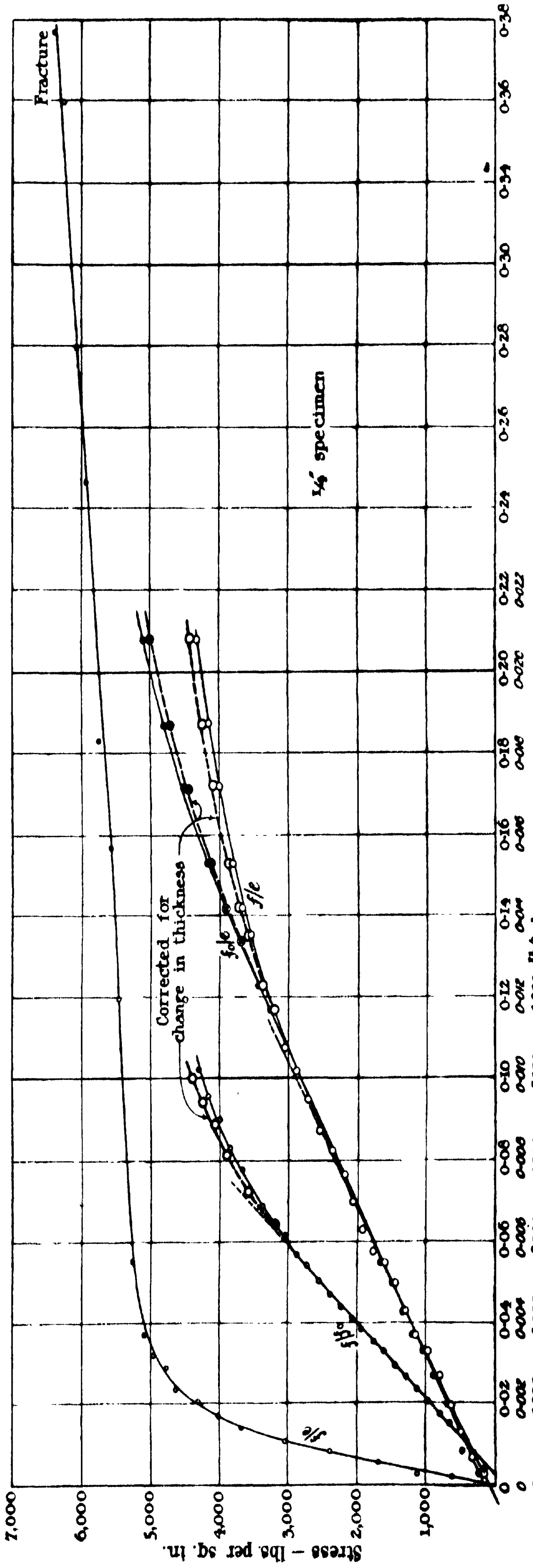

Stress – lbs. per sq. in.
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
0
0·02
0·04
0·06
0·08
0·10
0·12
0·14
0·16
0·18
0·20
0·22
0·24
0·26
0·28
0·30
0·34
0·36
0·38
lb/in²
Fracture
1/4" specimen
Corrected for change in thickness
f/e
f₀/e
f/f₀

Fig. 10.

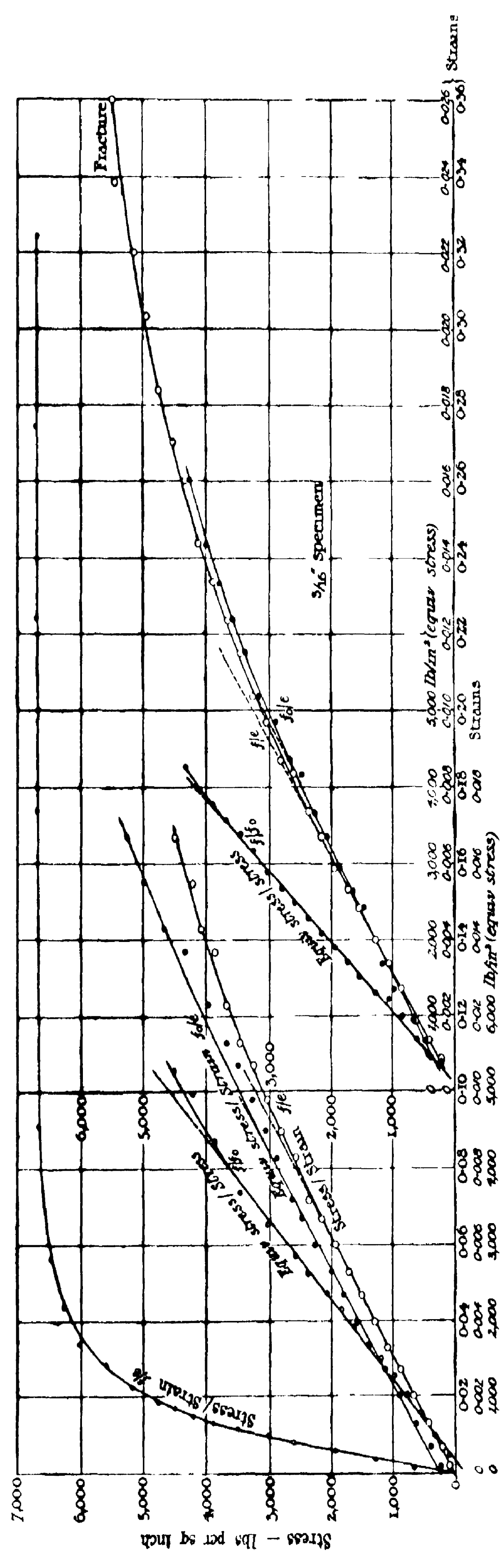

Stress – lbs per sq inch
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
Strains
0 0.02 0.04 0.06 0.08 0.10 0.12 0.14 0.16 0.18 0.20 0.22 0.24 0.26 0.28 0.30 0.32 0.34 0.36
0.002 0.004 0.006 0.008 0.010 0.012 0.014 0.016 0.018 0.020 0.022 0.024 0.026
1,000 2,000 3,000 4,000 5,000 6,000 lb/in² (equiv stress)
Fracture
3/16" specimen
Stress/Strain f/e
f/e
f₀/e
f/f₀
Equiv stress/Strain
Equiv stress/Stress

Fig 9

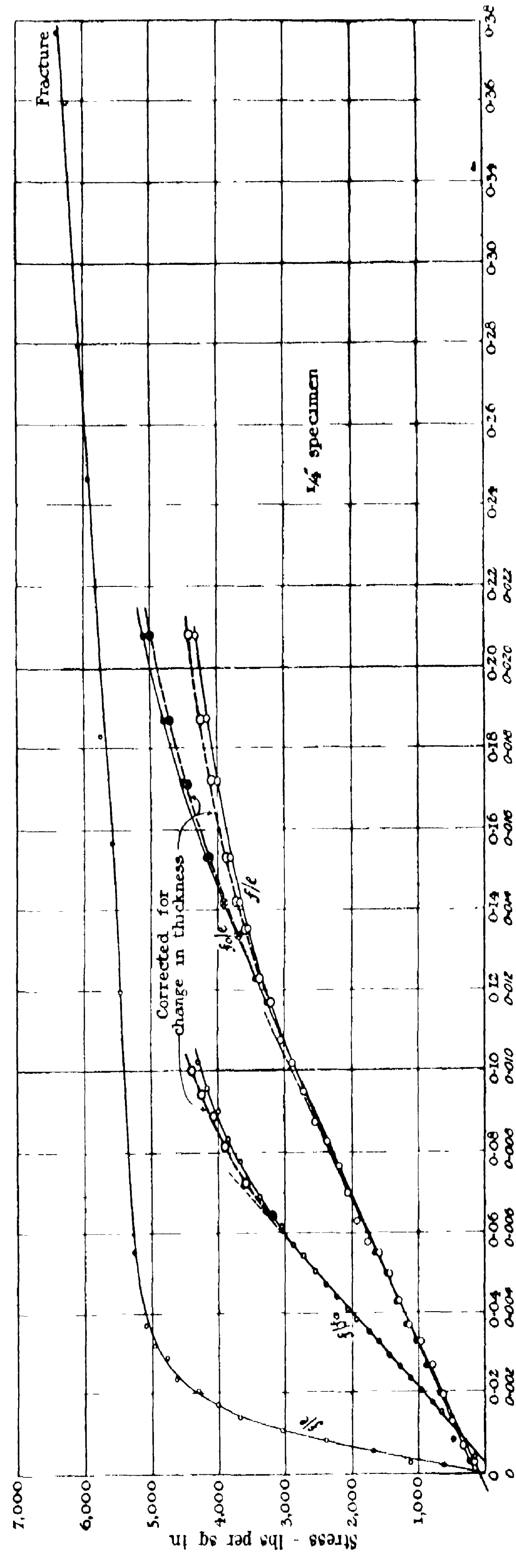

Stress - lbs per sq in
7,000
6,000
5,000
4,000
3,000
2,000
1,000
0
0 0.02 0.04 0.06 0.08 0.10 0.12 0.14 0.16 0.18 0.20 0.22 0.24 0.26 0.28 0.30 0.32 0.34 0.36 0.38
0.002 0.004 0.006 0.008 0.010 0.012 0.014 0.016 0.018 0.020 0.022
1,000 2,000 3,000 4,000 5,000 6,000 lb/in²
Fracture
1/4" specimen
Corrected for change in thickness
f/e
f₀/e
f/f₀

Fig 10

TABLE II.—Properties of the Specimens Tested.

Test length, 6-inch.

| Specimen | $\frac{1}{16}''$ (No. 1) | $\frac{1}{16}''$ (No. 2) | $\frac{1}{8}''$ | $\frac{5}{32}''$ | $\frac{3}{16}''$ | $\frac{1}{4}''$ | $\frac{3}{8}''$ |
|---|---|---|---|---|---|---|---|
| **Width** | 0·5086″ | 0·5090″ | 0·5050″ | 0·4992″ | 0·4978″ | 0·4972″ | 0·4992″ |
| **Thickness** | 0·0505″ | 0·0506″ | 0·1213″ | 0·1517″ | 0·1863″ | 0·2533″ | 0·4045″ |
| **Area (in.²)** | $0\cdot0257=\frac{1}{38\cdot85}\square$ | $0\cdot0258=\frac{1}{38\cdot75}\square''$ | $0\cdot0613\square''$ | $0\cdot07575=\frac{1}{13\cdot2}\square''$ | $0\cdot0928=\frac{1}{10\cdot76}\square''$ | $0\cdot1258=\frac{1}{7\cdot97}\square''$ | $0\cdot202=\frac{1}{4\cdot96}\square''$ |
| **Modulus** $\frac{\text{lb.}}{\text{in.}^2}$ | 362,000 | 355,000 | 295,000 | 324,000 | 313,000 | 309,000 | 251,000 |
| **POISSON'S ratio** | — | — | 2·7 | 2·5 | 2·6 | 2·3 | 2·4 |
| $\frac{C}{C_0}$ [optical] | 0·82 | 0·84 | 0·97 | 1·00 | 1·00 | 0·95 | 1·20 (?) |
| **Maximum extension before fracture** | 0·28″ | 1·66″ | 1·53″ | 2·14″ | 1·94″ | 2·14″ | 2·26″ |
| **Ultimate extension (after fracture)** | 0 | 1·23″ | 1·07″ | 1·67″ | 1·54″ | 1·72″ | 1·74″ |
| **Recovery in length** | 0·28″ | 0·43″ | 0·46″ | 0·47″ | 0·40″ | 0·42″ | 0·52″ |

The optical arrangements are modified for the new conditions as shown in the accompanying fig. 11. Light from the filament A of a Nernst lamp is focussed on to a vertical slit S by aid of a lens B, after passing through a Nicol's prism M and this narrow band is in turn focussed on the central section of the beam at D, and analysed by a second Nicol's prism N. A lens E placed at a convenient distance from the beam transmits this light as a parallel beam to a reflecting prism F, from which it passes through prisms G, H. The spectrum so obtained is focussed on a glass screen L ruled with lines $\frac{1}{100}$ inch apart, and provided with a micrometer eye-piece for measuring the ordinates of the bands observed.

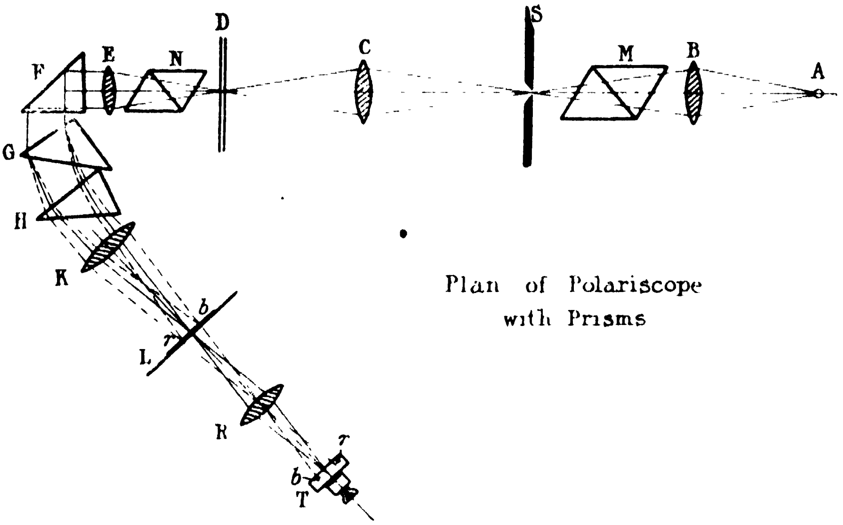


Fig. 11.

The field of view consists therefore of the spectrum of a Nernst lamp filament to which is added the effect produced by a narrow section of a beam of rectangular cross-section under pure bending moment. The relative retardation, owing to this latter stress effect, produces black bands in the field having a variable distance apart depending on the optical law of the retardation of the wave-length.

The general disposition of the field of view is shown in the accompanying fig. 12 in which bands of the first and second order appear on each side of the neutral axis C of the beam, and their co-ordinates are measured by reference to the graduations on the glass scale with the aid of a pair of parallel wires D, the positions of which can be adjusted vertically by a micrometer head E reading to $\frac{1}{6000}$ inch, while complete turns of the screw are obtained from a scale F on the left, which also appears in the field of view. In order to calibrate the horizontal scale the Nernst lamp and nitro-cellulose beam are removed, the Nicols rotated to parallelism, and a beam of solar light focussed on to the slit. The position of lines of known wave-length are noted with reference to the horizontal scale, and from these observations the constants in the equation

$$x_1 = A + \frac{B}{\lambda^2} + \frac{C}{\lambda^4}$$

are found for calibrating the positions of the black bands.

In the observations it is found that the depth of the beam does not appear quite constant throughout the field, an error due to the combined imperfections of the lenses and prisms employed. The maximum change of depth is about 2 per cent., and a correction is therefore necessary to reduce all vertical distances to a constant depth of beam.

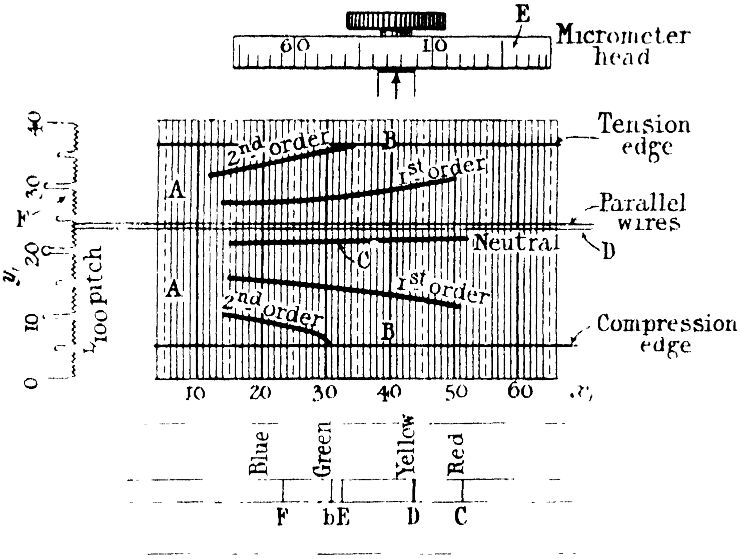


Fig. 12. View of spectrum.

The observations made are too numerous to give in detail, but typical examples of some measurements are shown in the accompanying fig. 13, in which the black bands, due to extinction of light, are drawn for a bending moment of 178·3 in pound and inch units. This, however, is not the exact appearance of the bands in the field of view owing to variation in the wave-length which alters the horizontal scale, but is here made uniform for plotting.

For some calculations, however, it is more convenient to show the form of the bands corresponding to a definite wave-length with a varying bending moment. Owing to the presence of a small amount of initial retardation in plates of nitro-cellulose, due to the method of manufacture, which leaves traces of initial stress, there is generally some slight difference between the bands on each side of the neutral axis, and a more accurate value is probable if the mean value for the two sides is taken.

If relative retardation is a linear function of the stress difference, these new abscissæ will represent the mean stress, but if the strain varies linearly they will also represent strains to another scale.

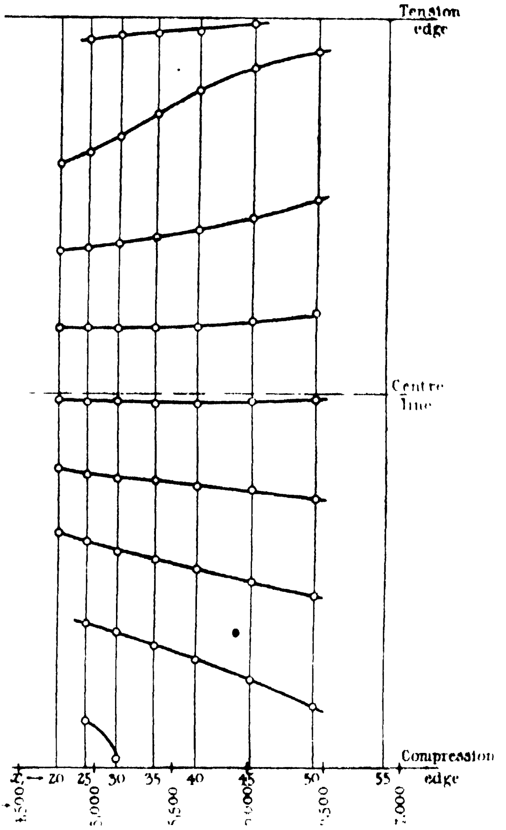


Fig. 13. Stress bands in spectrum load = 18 lb. B.mt. 178·3 lb. inches.

If then the mean distances of the bands are plotted as ordinates against the order of the band as abscissæ, fig. 14, a convenient form of diagram OCF is obtained in

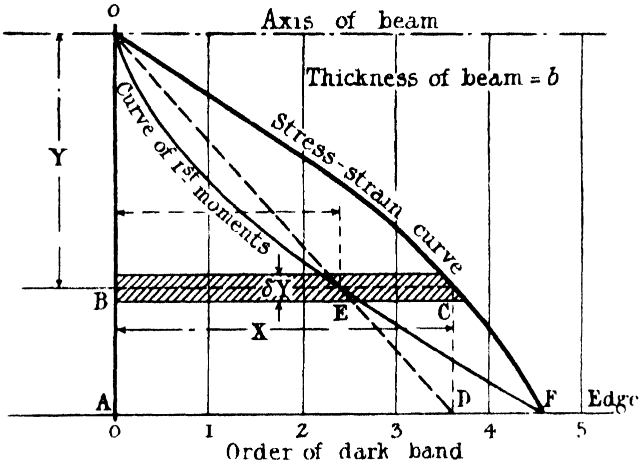


Fig. 14.

rectangular co-ordinates X, Y, in which Y is the distance from the neutral axis to a scale $\alpha$ and $f$ is the stress to a scale $\beta$, or

$$y = \mathrm{Y} \, . \, \alpha \, ; \qquad f = \mathrm{X} \, . \, \beta, \text{ say.}$$

Now the bending moment

$$M = \int_{-\frac{1}{2}d}^{+\frac{1}{2}d} fb \,.\, y \,.\, dy$$

for a breadth $= b$ and a depth $= d$, or

$$M = b \,.\, \alpha^2 \,.\, \beta \int_{-\frac{1}{2}d}^{+\frac{1}{2}d} X \,.\, Y \,.\, dY$$

$$= 2 \,.\, \alpha^2\beta \,.\, b \times \text{first moment of the area of the diagram about the neutral axis.}$$

If now any point C on the curve is projected on to the edge line at D and the line OD is drawn to the origin of co-ordinates intersecting the horizontal through C at E, then

$$BE : Y :: AD : AO = X : \frac{d}{2},$$

or

$$XY = BE \times \frac{d}{2}.$$

Hence $\int_{-\frac{1}{2}d}^{\frac{1}{2}d} X \,.\, Y \,.\, dY$ represents the area OEFAO $\times d$.

A typical example of one of these diagrams is shown in the accompanying fig. 15,

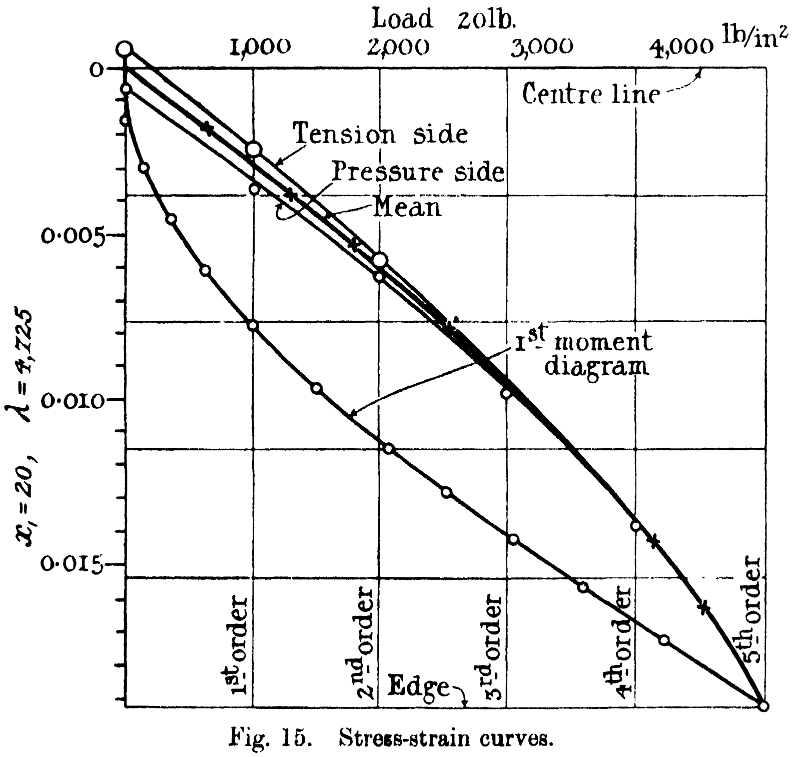


Fig. 15. Stress-strain curves.

of which about forty were actually prepared. The first moment areas M′, as determined by planimeter measurements, are shown in Table III., and when these are divided by the bending moment a value of $\alpha^2\beta b$ is obtained. If, however, the relative

TABLE III.—First Moments of Stress-strain Curves = M′.

| $x_1$. | λ. | W = 20 lb. | 18 lb. | 16 lb. | 14 lb. | 12 lb. | 10 lb. | 9 lb. |
|---|---|---|---|---|---|---|---|---|
| | | M = 198 lb./in. | 178·3 lb./in. | 158·7 lb./in. | 139·2 lb./in. | 119·6 lb./in. | 100·0 lb./in. | 90·2 lb./in. |
| 20 | 4725 | 46·20 | 40·30 | 33·50 | 29·65 | 25·60 | — | — |
| 25 | 4925 | 43·15 | 38·40 | 32·15 | 28·65 | 24·35 | — | — |
| 30 | 5135 | 41·25 | 36·50 | 31·80 | 27·00 | 23·35 | 19·95 | 17·90 |
| 35 | 5380 | 39·85 | 36·40 | 31·10 | 25·75 | 21·20 | 18·25 | — |
| 40 | 5660 | 38·00 | 34·65 | 28·65 | 24·00 | 20·55 | 17·50 | |
| 45 | 6015 | 36·65 | 31·90 | 27·85 | 23·00 | 19·25 | 17·00 | — |
| 50 | 6430 | 33·50 | 29·55 | 25·00 | 21·15 | 18·60 | 15·35 | — |

retardation is assumed to be independent of the wave-length, the mean value of $\frac{M'\lambda}{M}$ affords values of $\beta$ corresponding to different wave-lengths. The values of $\beta$ determined in this way are shown in the accompanying Table IV.

TABLE IV.

| $x_1$. | λ. | $\beta$. |
|---|---|---|
| 20 | 4725 | 920 |
| 25 | 4925 | 960 |
| 30 | 5135 | 1000 |
| 35 | 5380 | 1050 |
| 40 | 5660 | 1104 |
| 45 | 6015 | 1172 |
| 50 | 6430 | 1255 |

In order to determine the scale of strains the value of YOUNG'S modulus E = 309,000, as taken from the measurements in tension within the elastic limit, is assumed to hold near the neutral axis of the beam for all loads, and since in this region we have the strain $e = \text{Y} \,.\, s$, where $s$ is the scale for strains, then

$$\text{E} = \frac{df}{de} = \frac{\beta}{s} \cdot \frac{d\text{X}}{d\text{Y}},$$

or

$$s = \frac{\beta}{\text{E}} \cdot \frac{d\text{X}}{d\text{Y}}.$$

The slopes $\frac{d\text{X}}{d\text{Y}}$ near the neutral axis are measured from the diagrams similar to those of fig. 15, and their values are shown in the accompanying Table V.

multiplied by the values of $\beta$ appropriate to the wave-length. Their mean values afford measures of the strains as the table shows.

TABLE V.—Values of $\beta \cdot \frac{\delta X}{\delta Y}$.

| | | | | | | | |
|---|---|---|---|---|---|---|---|
| $x_1$ 20 | 1165 | 1012 | 805 | 681 | 589 | — | — |
| 25 | 1167 | 1000 | 810 | 672 | 581 | — | — |
| 30 | 1180 | 1000 | 810 | 670 | 572 | 473 | 435 |
| 35 | 1180 | 998 | 840 | 657 | 559 | 466 | — |
| 40 | 1196 | 1032 | 848 | 657 | 563 | 468 | — |
| 45 | 1201 | 996 | 838 | 656 | 563 | 479 | — |
| 50 | 1205 | 997 | 831 | 677 | 561 | 467 | — |
| Total . . . | 8294 | 7035 | 5782 | 4670 | 3991 | 2353 | 435 |
| Mean . . | 1185 | 1005 | 826 | 667 | 570 | 471 | 435 ? |
| Scale of strains $\frac{1}{E}\beta \cdot \frac{\delta X}{\delta Y}$ | $3 \cdot 84 \times 10^{-8}$ | $3 \cdot 25 \times 10^{-8}$ | $2 \cdot 67 \times 10^{-8}$ | $2 \cdot 16 \times 10^{-8}$ | $1 \cdot 85 \times 10^{-8}$ | $1 \cdot 52 \times 10^{-8}$ | $1 \cdot 39 \times 10^{-8}$ |
| $\frac{1}{s} \times 10^{-3}$ | 0·260″ | 0·308″ | 0·375″ | 0·463″ | 0·541″ | 0·658″ | 0·720″ |

The data afforded by this method is therefore sufficient to construct a stress-strain curve entirely from these measurements of the bands due to retardation in the spectrum, and if the assumptions are correct, it ought to agree with a similar diagram constructed from data obtained independently. The stress-strain diagram

TABLE VI.—Stress-strain Values.

| | $x_1$. | λ. | Strains. | | | | | | | | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | | 0·002. | 0·004. | 0·006. | 0·008. | 0·010. | 0·012. | 0·014. | 0·016. | 0·018. | 0·0192. |
| W = 20 lb. | 20 | 4725 | 596 | 1210 | 1815 | 2350 | 2834 | 3290 | 3700 | 4090 | 4420 | 4570 |
| | 25 | 4925 | 612 | 1205 | 1827 | 2362 | 2830 | 3290 | 3672 | 4015 | 4340 | 4530 |
| | 30 | 5135 | 620 | 1230 | 1840 | 2360 | 2810 | 3250 | 3620 | 3950 | 4220 | 4340 |
| | 35 | 5380 | 629 | 1258 | 1855 | 2390 | 2850 | 3280 | 3644 | 3980 | 4270 | 4400 |
| | 40 | 5660 | 631 | 1262 | 1880 | 2424 | 2900 | 3330 | 3730 | 4020 | 4265 | 4360 |
| | 45 | 6015 | 647 | 1271 | 1860 | 2420 | 2910 | 3255 | 3765 | 4130 | 4450 | 4610 |
| | 50 | 6430 | 654 | 1295 | 1910 | 2440 | 2880 | 3295 | 3695 | 4000 | 4260 | 4370 |
| Mean . . . | | | 627 | 1247 | 1855 | 2391 | 2859 | 3284 | 3689 | 4026 | 4311 | 4454 |

obtained from spectrum observations gives the following values, of which Table VI. is a typical example, from which the mean values are obtained as follows:—

| Strains . . . . . . | 0·001 | 0·002 | 0·003 | 0·004 | 0·005 | 0·006 | 0·007 | 0·008 |
|---|---|---|---|---|---|---|---|---|
| Stresses in lbs. per sq. inch . . . . . . | 310 | 620 | 920 | 1240 | 1540 | 1845 | 2135 | 2400 |
| Strains . . . . . . | 0·009 | 0·010 | 0·011 | 0·012 | 0·013 | 0·014 | 0·018 | 0·0192 |
| Stresses in lbs. per sq. inch . . . . . . | 2665 | 2890 | 3055 | 3285 | 3460 | 3680 | 4310 | 4455 |

Table of mean values

| Strain | 0·001 | 0·002 | 0·003 | 0·004 | 0·005 | 0·006 | 0·007 | 0·008 | 0·009 | 0·010 | 0·011 | 0·012 | 0·013 | 0·014 | 0·016 | 0·018 | 0·0192 |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| Stress | 310 | 620 | 918 | 1,238 | 1,541 | 1,844 | 2,135 | 2,399 | 2,665 | 2,889 | 3,056 | 3,283 | 3,459 | 3,682 | 3,997 | 4,311 | 4,454 |
| Corrected for thickness | 310 | 621 | 919 | 1,240 | 1,544 | 1,849 | 2,141 | 2,407 | 2,675 | 2,901 | 3,070 | 3,300 | 3,478 | 3,703 | 4,023 | 4,343 | 4,490 |

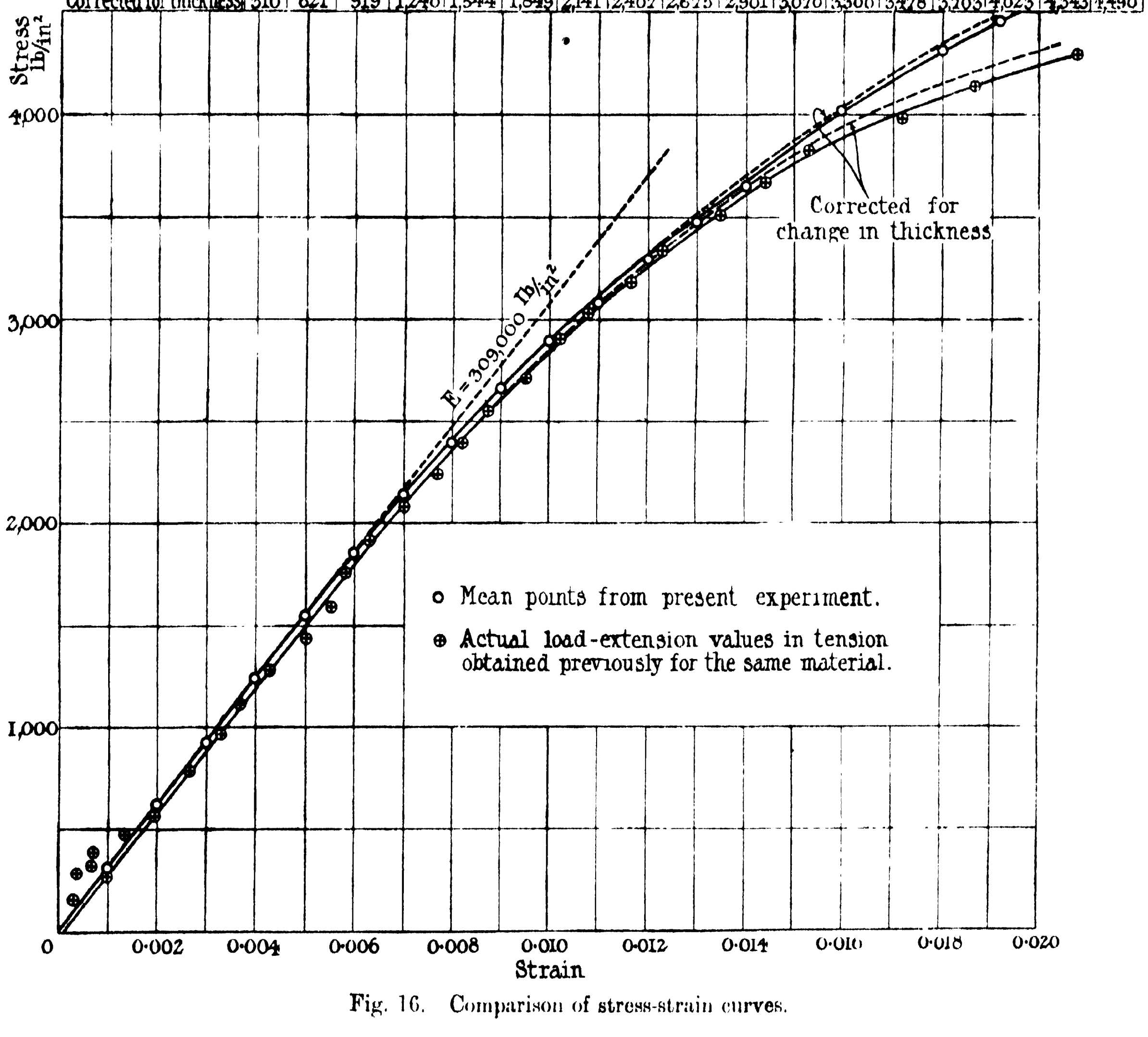


Fig. 16. Comparison of stress-strain curves.

This information has already been obtained however by observations on precisely similar material under direct tension stress, Table II. and fig. 13, and on comparing plots of the two sets of data obtained, fig. 16, the agreement is seen to be a remarkably close one up to about 3500 lbs. per sq. inch. This agreement is improved if the changes in thickness are allowed for, since the corrected curves then lie closer together, and strengthen the evidence in favour of the law of optical retardation being an effect of stress and not of strain, and also that it is still a linear function much beyond the elastic limit of the material.

The whole of the evidence, in fact, appears to show that the transparent nitro-cellulose examined obeys a linear stress optical law which holds up to approximately twice the range of the elastic limit of stress; and that within this range optical determinations of stress distribution may be relied upon.

In conclusion we desire to express our grateful thanks for the help afforded in this work by the Department of Scientific and Industrial Research, also for valuable suggestions from Prof. FILON, F.R.S., and Prof. PORTER, F.R.S., during its progress, and for the skilful assistance of Mr. F. H. WITHYCOMBE in preparing all the experimental apparatus required.

# VI. *The Phenomena of Rupture and Flow in Solids.*

*By* A. A. GRIFFITH, *M. Eng.* (*of the Royal Aircraft Establishment*).

*Communicated by* G. I. TAYLOR, *F.R.S.*

Received February 11,— Read February 26, 1920.

## CONTENTS.

| | | Page |
|---|---|---|
| 1. | Introduction | 163 |
| 2. | A Theoretical Criterion of Rupture | 165 |
| 3. | Application of the Theory to a Cracked Plate | 167 |
| 4. | Experimental Verification of the Theory | 172 |
| 5. | Deductions from the Foregoing Results | 177 |
| 6. | The Strength of Thin Fibres | 180 |
| 7. | Molecular Theory of Strength Phenomena | 183 |
| 8. | Extended Application of the Molecular Orientation Theory | 186 |
| 9. | Practical Limitations of the Elastic Theory | 192 |
| 10. | Methods of Increasing the Strength of Materials | 194 |
| 11. | Application of the Theory to Liquids | 195 |
| 12. | Summary of Conclusions | 197 |

## 1. *Introduction.*

IN the course of an investigation of the effect of surface scratches on the mechanical strength of solids, some general conclusions were reached which appear to have a direct bearing on the problem of rupture, from an engineering standpoint, and also on the larger question of the nature of intermolecular cohesion.

The original object of the work, which was carried out at the Royal Aircraft Establishment, was the discovery of the effect of surface treatment—such as, for instance, filing, grinding or polishing—on the strength of metallic machine parts subjected to alternating or repeated loads. In the case of steel, and some other metals in common use, the results of fatigue tests indicated that the range of alternating stress which could be permanently sustained by the material was smaller than the range within which it was sensibly elastic, after being subjected to a great number of reversals. Hence it was inferred that the safe range of loading of a part, having a scratched or

[Published October 21, 1920.

grooved surface of a given type, should be capable of estimation with the help of one of the two hypotheses of rupture commonly used for solids which are elastic to fracture. According to these hypotheses rupture may be expected if (*a*) the maximum tensile stress, (*b*) the maximum extension, exceeds a certain critical value. Moreover, as the behaviour of the materials under consideration, within the safe range of alternating stress, shows very little departure from HOOKE's law, it was thought that the necessary stress and strain calculations could be performed by means of the mathematical theory of elasticity.

The stresses and strains due to typical scratches were calculated with the help of the mathematical work of Prof. C. E. INGLIS,* and the soap-film method of stress estimation developed by Mr. G. I. TAYLOR in collaboration with the present author.† The general conclusions were that the scratches ordinarily met with could increase the maximum stresses and strains from two to six times, according to their shape and the nature of the stresses, and that these maximum stresses and strains were to all intents and purposes independent of the absolute size of the scratches. Thus, on the maximum tension hypothesis, the weakening of, say, a shaft 1 inch in diameter, due to a scratch one ten-thousandth of an inch deep, should be almost exactly the same as that due to a groove of the same shape one-hundredth of an inch deep.

These conclusions are, of course, in direct conflict with the results of alternating stress tests. So far as the author is aware, the greatest weakening due to surface treatment, recorded in published work, is that given by J. B. KOMMERS,‡ who found that polished specimens showed an increased resistance over turned specimens of 45 to 50 per cent. The great majority of published results indicate a diminution in strength of less than 20 per cent. Moreover, it is certain that reducing the size of the scratches increases the strength.

To explain these discrepancies, but one alternative seemed open. Either the ordinary hypotheses of rupture could be at fault to the extent of 200 or 300 per cent., or the methods used to compute the stresses in the scratches were defective in a like degree.

The latter possibility was tested by direct experiment. A specimen of soft iron wire, about 0·028-inch diameter and 100 inches long, which had a remarkably definite elastic limit, was selected. This was scratched spirally (*i.e.*, the scratches made an angle of about 45 degrees with the axis) with carborundum cloth and oil. It was then normalised to remove initial stresses and subjected to a tensile load. Under these conditions the effect of the spiral scratches was to impart a twist to the wire, the twisting couple arising entirely from the stress-system due to the scratches. It was found that if the load exceeded a certain critical value, a part of the twist, amounting in some cases to 15 per cent., remained after the removal of the load. It was inferred that at this critical load the maximum stresses in the scratches reached the elastic limit of the material. This load was about one-quarter to one-third of that which caused the wire to yield as a whole, so that the scratches increased the maximum stress three or four times. The readings were quite definite even in the case of scratches produced by No. 0 cloth, which were found by micrographic examination to be but $10^{-4}$-inch deep. Control experiments with longitudinal and circumferential scratches gave twists only 2 or 3 per cent. of those found with spiral scratches, and there was no permanent twist.

---

* "Stresses in a Plate due to the Presence of Cracks and Sharp Corners," 'Proc. Inst. Naval Architects,' March 14, 1913.

† 'Proc. Inst. Mech. Eng.' December 14, 1917, pp. 755–809.

‡ 'Intern. Assoc. for Testing Materials,' 1912, vol. 4A and vol. 4B.

This substantial confirmation of the estimated values of the stresses, even in very fine scratches, shows that the ordinary hypotheses of rupture, as usually interpreted, are inapplicable to the present phenomena. Apart altogether from the numerical discrepancy, the observed difference in fatigue strength as between small and large scratches presents a fundamental difficulty.

## 2. *A Theoretical Criterion of Rupture.*

In view of the inadequacy of the ordinary hypotheses, the problem of the rupture of elastic solids has been attacked from a new standpoint. According to the well-known "theorem of minimum energy," the equilibrium state of an elastic solid body, deformed by specified surface forces, is such that the potential energy of the whole system* is a minimum. The new criterion of rupture is obtained by adding to this theorem the statement that the equilibrium position, if equilibrium is possible, must be one in which rupture of the solid has occurred, if the system can pass from the unbroken to the broken condition by a process involving a continuous decrease in potential energy.

In order, however, to apply this extended theorem to the problem of finding the breaking loads of real solids, it is necessary to take account of the increase in potential energy which occurs in the formation of new surfaces in the interior of such solids. It is known that, in the formation of a crack in a body composed of molecules which attract one another, work must be done against the cohesive forces of the molecules on either side of the crack.† This work appears as potential surface energy, and if the width of the crack is greater than the very small distance called the "radius of molecular action," the energy per unit area is a constant of the material, namely, its surface tension.

In general, the surfaces of a small newly formed crack cannot be at a distance apart greater than the radius of molecular action. It follows that the extended theorem of minimum energy cannot be applied unless the law connecting surface energy with distance of separation is known.

---

* POYNTING and THOMSON, 'Properties of Matter,' ch. xv.

† The potential energy of the applied surface forces is, of course, included in the "potential energy, of the system."

There is, however, an important exception to this statement. If the body is such that a crack forms part of its surface in the unstrained state, it is not to be expected that the spreading of the crack, under a load sufficient to cause rupture, will result in any large change in the shape of its extremities. If, further, the crack is of such a size that its width is greater than the radius of molecular action at all points except very near its ends, it may be inferred that the increase of surface energy, due to the spreading of the crack, will be given with sufficient accuracy by the product of the increment of surface into the surface tension of the material.

The molecular attractions across such a crack must be small except very near its ends; it may therefore be said that the application of the mathematical theory of elasticity on the basis that the crack is assumed to be a traction-free surface, must give the stresses correctly at all points of the body, with the exception of those near the ends of the crack. In a sufficiently large crack the error in the strain energy so calculated must be negligible. Subject to the validity of the other assumptions involved, the strength of smaller cracks calculated on this basis must evidently be too low.

The calculation of the potential energy is facilitated by the use of a general theorem which may be stated thus: In an elastic solid body deformed by specified forces applied at its surface, the sum of the potential energy of the applied forces and the strain energy of the body is diminished or unaltered by the introduction of a crack whose surfaces are traction-free.

This theorem may be proved* as follows: It may be supposed, for the present purpose, that the crack is formed by the sudden annihilation of the tractions acting on its surface. At the instant following this operation, the strains, and therefore the potential energy under consideration, have their original values; but, in general, the new state is not one of equilibrium. If it is not a state of equilibrium, then, by the theorem of minimum energy, the potential energy is reduced by the attainment of equilibrium; if it is a state of equilibrium the energy does not change. Hence the theorem is proved.

Up to this point the theory is quite general, no assumption having been introduced regarding the isotropy or homogeneity of the substance, or the linearity of its stress-strain relations. It is necessary, of course, for the strains to be elastic. Further progress in detail, however, can only be made by introducing HOOKE's law.

If a body having linear stress-strain relations be deformed from the unstrained state to equilibrium by given (constant) surface forces, the potential energy of the latter is diminished by an amount equal to twice the strain energy.† It follows that the net reduction in potential energy is equal to the strain energy, and hence the total decrease in potential energy due to the formation of a crack is equal to the increase in strain energy less the increase in surface energy. The theorem proved above shows that the former quantity must be positive.

---

* The proof is due to Mr. C. WIGLEY, late of the Royal Aircraft Establishment.

† A. E. H. LOVE, 'Mathematical Theory of Elasticity,' 2nd ed., p. 170.

### 3. *Application of the Theory to a Cracked Plate.*

The necessary analysis may be performed in the case of a flat homogeneous isotropic plate of uniform thickness, containing a straight crack which passes normally through it, the plate being subjected to stresses applied in its plane at its outer edge.

If the plate is thin, its state is one of "plane stress," and in this case it may, without additional complexity, be subjected to any uniform stress normal to its surface, in addition to the edge tractions. If it is not thin, it may still be dealt with provided it is subjected to normal surface stresses so adjusted as to make the normal displacement zero. Here the plate is in a state of "plane strain." The equations to the two states are of the same form,* differing only in the value of the constants; they will therefore be taken together.

The strain energy may be found, with sufficient accuracy, in the general case where the edge-tractions are arbitrary; it is necessary in the present application, however, for the resulting stress-system to be symmetrical about the crack, as otherwise it is not obvious that the latter will remain straight as it spreads. The only stress distribution which will be considered, therefore, is that in which the principal stresses in the plane of the plate, at points far from the crack, are respectively parallel and perpendicular to the crack, and are the same at all such points. This is equivalent to saying that, in the absence of the crack, the plate would have been subjected to uniform principal stresses in and perpendicular to its plane. It is also necessary, on physical grounds, for the stress perpendicular to the crack and in the plane of the plate to be a tension, otherwise the surfaces of the crack are forced together instead of being separated, and they cannot remain free from traction.

In calculating the strain energy of the plate use will be made of the solution obtained by Prof. INGLIS for the stresses in a cracked plate, to which reference has already been made. The notation of Prof. INGLIS's paper will be employed. In that notation $\alpha$, $\beta$, are elliptic co-ordinates defined by the family of confocal ellipses; $\alpha =$ const. and the orthogonal family of hyperbolæ $\beta =$ const. The crack is represented by the limiting ellipse or focal line $\alpha = 0$. The axis of $x$ coincides with the major axes, and the axis of $y$ with the minor axes of the ellipses. The cartesian co-ordinates $x$, $y$, are connected with the elliptic co-ordinates $\alpha$, $\beta$, by the relation

$$x + iy = c \cosh (\alpha + i\beta).$$

$R_{\alpha\alpha}$, $u_\alpha$, are the tensile stress and displacement respectively along the normal to $\alpha =$ const.

$R_{\beta\beta}$, $u_\beta$, are the corresponding quantities in the case of the normal to $\beta =$ const.

$S_{\alpha\beta}$ is the shear stress in the directions of these normals.

$c$ is the half-length of the focal line.

---

* A. E. H. LOVE, 'Mathematical Theory of Elasticity,' 2nd ed., p. 205.

$h$ is the modulus of transformation, $\sqrt{\dfrac{2}{c^2\,(\cosh 2\alpha - \cos 2\beta)}}$.

$\mu$ is the modulus of rigidity of the material.

E is YOUNG'S modulus.

$\sigma$ is POISSON'S ratio.

$p = 3 - 4\,\sigma$ in the case of plane strain, and

$\dfrac{3-\sigma}{1+\sigma}$ in the case of plane stress.

The state of uniform stress existing at points far from the crack (*i.e.* where $\alpha$ is large) will be specified by the three principal tensions P, Q and R. P is normal to the plate, and in the case of plane stress it is the same everywhere. Q and R are parallel respectively to the axes of $x$ and $y$, and R is positive.

The strain energy of the plate is a quadratic function of P, Q and R, and hence, in accordance with the theorem proved above, the increase of strain energy due to the crack must be a positive quadratic function of P, Q and R. The general form of this function may be found by evaluating a sufficient number of particular cases.

The following particular cases are sufficient :—

I.—Q = R (and P = 0 in the case of plane stress).

Boundary of crack given by $\alpha = \alpha_0$.

The stresses are

$$R_{\alpha\alpha} = R\,\frac{\sinh 2\alpha\,(\cosh 2\alpha - \cosh 2\alpha_0)}{(\cosh 2\alpha - \cos 2\beta)^2}, \quad \ldots \ldots \ldots \ldots \quad (1)$$

$$R_{\beta\beta} = R\,\frac{\sinh 2\alpha\,(\cosh 2\alpha + \cosh 2\alpha_0 - 2\cos 2\beta)}{(\cosh 2\alpha - \cos 2\beta)^2}, \quad \ldots \ldots \quad (2)$$

$$S_{\alpha\beta} = R\,\frac{\sin 2\beta\,(\cosh 2\alpha - \cosh 2\alpha_0)}{(\cosh 2\alpha - \cos 2\beta)^2}, \quad \ldots \ldots \ldots \ldots \quad (3)$$

while the displacements are given by

$$\left.\begin{aligned} \frac{u_\alpha}{h} &= \frac{c^2 R}{8\mu}\{(p-1)\cosh 2\alpha - (p+1)\cos 2\beta + 2\cosh 2\alpha_0\} \\ \frac{u_\beta}{h} &= 0 \end{aligned}\right\} \quad \ldots \quad (4)$$

The strain energy of the material within the ellipse $\alpha$, per unit thickness of plate is

$$\tfrac{1}{2}\int_0^{2\pi} \frac{u_\alpha}{h}\,.\,R_{\alpha\alpha}\,.\,d\beta + \tfrac{1}{2}\int_0^{2\pi} \frac{u_\beta}{h}\,.\,S_{\alpha\beta}\,.\,d\beta. \quad \ldots \ldots \ldots \quad (5)$$

On substituting and integrating, it is found that, as $\alpha$ becomes large, the strain energy tends towards the value

$$\frac{\pi c^2 R^2}{8\mu}\left\{\tfrac{1}{2}(p-1)\,e^{2\alpha}+(3-p)\cosh 2_{\alpha_0}\right\} \quad \text{. . . . . . . .} \quad (6)$$

Hence W, the increase of strain energy due to the cavity $\alpha_0$, is given by

$$W = \frac{\pi c^2 R^2}{8\mu}(3-p)\cosh 2\alpha_0 \quad \text{. . . . . . . . . .} \quad (7)$$

or, on proceeding to the limit, $\alpha_0 = 0$,

$$W = \frac{(3-p)\,\pi c^2 R^2}{8\mu} \quad \text{. . . . . . . . . . . .} \quad (8)$$

for a very narrow crack of length $2c$.

II.—$R = 0$ ($= P$ in the case of plane stress) $\alpha_0 = 0$.

Here the stresses are entirely unaltered by the crack, at every point of the plate except the two points $x = \pm c$, $y = 0$, where $R\alpha\alpha = -Q$. It follows that $W = 0$.

III.—$Q = R = 0$, $\alpha_0 = 0$.

Here, again, the stresses are unaltered, and $W = 0$.

The only positive quadratic function of P, Q and R which is compatible with these three particular cases is that given by equation (8); this is therefore the general form of W, and rupture is determined entirely by the stress R, perpendicular to the crack.

A point of some interest, with regard to equation (8), may be noticed in passing. Since W cannot be negative it follows that, in real substances, where $\mu$ is positive, $3 - p$ must be positive. Hence $\sigma$ cannot be negative in real isotropic solids.

The potential energy of the surface of the crack, per unit thickness of the plate is

$$U = 4cT \quad \text{. . . . . . . . . . . . .} \quad (9)$$

where T is the surface tension of the material.

Hence the total diminution of the potential energy of the system, due to the presence of the crack, is

$$W - U = \frac{(3-p)\,\pi c^2 R^2}{8\mu} - 4cT. \quad \text{. . . . . . .} \quad (10)$$

The condition that the crack may extend is

$$\frac{\partial}{\partial c}(W-U) = 0,$$

or

$$(3-p)\,\pi c R^2 = 16\mu T, \quad \text{. . . . . . . . . .} \quad (11)$$

so that the breaking stress is

$$R = 2\sqrt{\frac{\mu T}{\pi\sigma c}}, \qquad (12)$$

in the case of plane strain, and

$$R = \sqrt{\frac{2ET}{\pi\sigma c}} \qquad (13)$$

in the case of plane stress.

Formula (13) has been verified experimentally. In connection with the experiments, interest attaches not only to the magnitude of R, but also to the value of the maximum tension in the material, which occurs at the extremities of the crack. This stress may be estimated if the radius of curvature of the boundary of the crack, at the points in question, can be found.

Expression (2) gives the maximum tension as

$$2R\sqrt{\frac{c}{\rho}} \qquad (14)$$

in case I. above, $\rho$ being the radius of curvature at the corners of the elliptic crack. Prof. INGLIS shows that this expression may also be used, with little error, for cracks which are elliptic only near their ends. The foregoing expressions for the stresses are obtained, however, on the assumption that the displacements are everywhere so small that their squares may be neglected. At the corner of a very sharp crack, it cannot be assumed, without proof, that the change in $\rho$ leaves formula (14) substantially unaffected.

In the case under consideration the displacements at the surface of the crack, due to a small tension $dR$ at distant points, are given by

$$\left.\begin{aligned} \frac{u_\alpha}{h} &= \frac{c^2 dR}{E}(\cosh 2\alpha_0 - \cos 2\beta) \\ \frac{u_\beta}{h} &= 0 \end{aligned}\right\} \qquad (15)$$

Whence, by resolution, the displacements parallel respectively to the major and minor axes are

$$\left.\begin{aligned} u_x &= \frac{2dR}{E} c \sinh \alpha_0 \cos \beta \\ u_y &= \frac{2dR}{E} c \cosh \alpha_0 \sin \beta \end{aligned}\right\}, \qquad (16)$$

which may be written

$$\left.\begin{aligned} u_x &= \frac{2dR}{E} x \tanh \alpha_0 \\ u_y &= \frac{2dR}{E} y \coth \alpha_0 \end{aligned}\right\} \qquad (17)$$

Equations (17) show that the effect of the small stress $d\mathrm{R}$ on the elliptic cavity is to deform it into another ellipse. If $a$ and $b$ are the major and minor semi-axes of the ellipse, when the plate is subjected to a stress R, then, by (17),

$$\left.\begin{aligned}\frac{da}{d\mathrm{R}} &= \frac{2b}{\mathrm{E}}\\ \frac{db}{d\mathrm{R}} &= \frac{2a}{\mathrm{E}}\end{aligned}\right\} \quad \ldots \ldots \ldots \quad (18)$$

on making use of the relation $b = a \tanh \alpha_0$.

The solution of these simultaneous differential equations is

$$\left.\begin{aligned}a &= a_0 \cosh \frac{2\mathrm{R}}{\mathrm{E}} + b_0 \sinh \frac{2\mathrm{R}}{\mathrm{E}}\\ b &= a_0 \sinh \frac{2\mathrm{R}}{\mathrm{E}} + b_0 \cosh \frac{2\mathrm{R}}{\mathrm{E}}\end{aligned}\right\} \quad \ldots \ldots \ldots \quad (19)$$

where $a_0$ and $b_0$ are the values of $a$ and $b$ in the unstrained state.

With the help of equations (19) it is possible to find the maximum stress, F, due to an applied stress, R, taking account of the change in the shape of the cavity. From (2)

$$\frac{d\mathrm{F}}{d\mathrm{R}} = 2\frac{a}{b} \quad \ldots \ldots \ldots \quad (20)$$

whence

$$\mathrm{F} = 2\int_0^{\mathrm{R}} \frac{a_0 \cosh \frac{2\mathrm{R}}{\mathrm{E}} + b_0 \sinh \frac{2\mathrm{R}}{\mathrm{E}}}{a_0 \sinh \frac{2\mathrm{R}}{\mathrm{E}} + b_0 \cosh \frac{2\mathrm{R}}{\mathrm{E}}}\, d\mathrm{R}$$

$$= \mathrm{E} \log \left(\cosh \frac{2\mathrm{R}}{\mathrm{E}} + \frac{a_0}{b_0} \sinh \frac{2\mathrm{R}}{\mathrm{E}}\right), \quad \ldots \ldots \quad (21)$$

and in the case of a narrow crack which is elliptic only near its ends, $\frac{a_0}{b_0}$ may, as in (14), be replaced by $\sqrt{\frac{c}{\rho}}$.

In the general case, where Q is not equal to R, the quantity R — Q must be added to the value of F given by (21).

Formulæ (19) and (21) are not, of course, exactly true. The application of integration to equations (18) and (20) involves the assumption that the strains are so small that they can be superposed. If the strains are finite, this involves an error in the stresses depending on the square of the strains. In the case of ordinary solids, it is improbable that this assumption can alter the calculated stress by as much as 1 per cent.

### 4. *Experimental Verification of the Theory.*

In order to test formula (13), it was necessary to select an isotropic material which obeyed HOOKE's law somewhat closely at all stresses, and whose surface tension at ordinary temperatures could be estimated. For these reasons glass was preferred to the metals in common use. A comparatively hard English glass,* having the following properties, was employed :—

Composition—$SiO_2$, 69·2 per cent. ; $K_2O$, 12·0 per cent. ; $Na_2O$, 0·9 per cent. ; $Al_2O_3$, 11·8 per cent. ; CaO, 4·5 per cent. ; MnO, 0·9 per cent.
Specific gravity—2·40.
YOUNG's modulus—9·01 × $10^6$ lbs. per sq. inch.
POISSON's ratio—0·251.
Tensile strength—24,900 lbs. per sq. inch.

The three last-named quantities were determined by the usual tension and torsion tests on round rods or fibres about 0·04-inch diameter and 3 inches long between the gauge points. The fibres had enlarged spherical ends which were fixed into holders with sealing wax. A slight load was applied while the wax was still soft, to ensure freedom from bending. The possible error of the extension measurements was about ± 0·3 per cent., and HOOKE's law was obeyed to this order of accuracy. No "elastic after-working" was observed with this glass, though more accurate measurements would doubtless have indicated its existence.

The problem of estimating the surface tension of glass, in the solid state, evidently requires special consideration. Direct determinations appeared to be impracticable, and ultimately an indirect method was decided on, in which the surface tension was found at a number of high temperatures and the value at ordinary temperatures deduced by extrapolation.

On the accepted theory of matter, intermolecular forces in solids and liquids consist mainly of two parts, namely, an attraction which increases rapidly as the distance between the molecules diminishes, balanced by a repulsion (the intrinsic pressure), which is due to the thermal vibrations of the molecules. It is reasonable to assume that the attraction, at constant volume, is sensibly independent of the temperature ; this amounts merely to supposing that the attraction exerted by a molecule does not depend on its state of motion. On this view, the temperature variation, at constant volume, of the intermolecular forces is determined entirely by the change in thermal energy. Hence, it may be inferred, on the accepted theory of surface tension,† that the surface tension of a material, at constant volume, is equal to a constant diminished by a quantity proportional to the thermal energy of the substance. In the case of solids, nearly the same result should hold at constant pressure, as the temperature-volume change is small.

---

* Supplied in the form of test-tubes by Messrs. J. J. Griffin, Kingsway, London.

† POYNTING and THOMSON, 'Properties of Matter,' ch. xv.

The specific heat of glass is greater at high than at low temperatures, but the temperature coefficient is not large. Hence its surface tension may be expected to be nearly a linear function of the temperature, and extrapolation should be fairly reliable. This was found to be the case with the glass selected for the present experiments.

In the neighbourhood of 1100° C. the surface tension was found by QUINCKE's drop method. At lower temperatures this method was not satisfactory, on account of the large viscosity of the liquid glass; but between 730° C. and 900° C. the method described below was found to be practicable. Fibres of glass, about 2 inches long and from 0·002-inch to 0·01-inch diameter, with enlarged spherical ends, were prepared. These were supported horizontally in stout wire hooks and suitable weights were hung on their mid-points. The enlarged ends prevented any sagging except that due to extension of the fibres. The whole was placed in an electric resistance furnace maintained at the desired temperature. Under these conditions viscous stretching of the fibre occurred until the suspended weight was just balanced by the vertical components of the tension in the fibre. The latter was entirely due, in the steady state, to the surface tension of the glass, whose value could therefore be calculated from the observed sag of the fibre. In the experiments the angle of sag was observed through a window in the furnace by means of a telescope with a rotating cross wire. If $w$ is the suspended weight, $d$ the diameter of the fibre, T the surface tension, and $\theta$ the angle at the point of suspension between the two halves of the fibre, then, evidently,

$$\pi \,.\, d \,.\, \mathrm{T} \,.\, \sin \tfrac{1}{2}\theta = w.$$

For this method of determining the surface tension to be valid, it is evidently necessary that the angle of sag shall reach a steady value before the development of local contractions, arising from the instability of liquid cylinders, becomes appreciable. That this requirement is satisfied is shown by the following experimental results. After heating for two hours at about 750° C. the angle of sag of a particular fibre was 18°·25. Two hours later it had increased by less than 0°·1. The temperature was then raised momentarily to 940° C., and quickly reduced again to 750° C. The angle was then found to be 20°·2. After two hours further heating at 750° C. the angle had *decreased* to 18°·4, agreeing within permissible limits of error with the former value. That is to say, substantially the same limiting angle of sag was reached whether the initial angle was above or below that limit.

Above 900° C. it was found that the viscosity was insufficient to enable an observation to be made before the fibre commenced to break up into globules. Below 730° C., on the other hand, observations made on fibres of different diameters were inconsistent, the apparent surface tension being higher for the larger fibres. The obvious meaning of this result is that below 730° C. the glass used was not a perfect viscous liquid and hence the method was inapplicable. The transition from the viscous liquid state was quite gradual. The maximum tension (apart from surface tension) which could be permanently sustained, was zero at 730° C., 1·3 lbs. per sq. inch at 657° C., and 24 lbs.

per sq. inch at 540° C. At lower temperatures the rates of increase, both of this "solid stress" and the viscosity, were enormously greater. At 540° C. a fibre took about 70 hours to reach the steady state.

Table I. below gives the values of the surface tension obtained from these experiments. That for the temperature 1110° C. is the mean of five determinations by the drop method. The remaining figures were obtained from the sag of fibres.

TABLE I.—Surface Tension of Glass.

| Temperature. | Surface Tension. |
|---|---|
| ° C. | lb. per inch. |
| 1110 | 0·00230 |
| 905 | 0·00239 |
| 896 | 0·00250 |
| 852 | 0·00249 |
| 833 | 0·00254 |
| 820 | 0·00249 |
| 801 | 0·00257 |
| 760 | 0·00255 |
| 745 | 0·00251 |
| 15 | 0·0031* |

So far as they go, these figures confirm the deduction that the surface tension of glass is approximately a linear function of temperature. Moreover, as the actual variation is not great, the error involved in assuming such a law and extrapolating to 15° C. is doubtless fairly small. The value so obtained, 0·0031 lb. per inch, will be used in the present application.

Rigorously, expressions (13) and (21) above are true only for small cracks in large flat plates. In view, however, of the difficulties attendant on annealing and loading large flat glass plates, it was decided to perform the breaking tests on thin round tubes and spherical bulbs. These were cracked and then annealed and broken by internal pressure. The calculation cannot be exact for such bodies, but the error may obviously be reduced by increasing the ratio of the diameter of the bulb or tube to the length of the crack. It will be seen from the results of the tests that the variation of this ratio from two to ten caused little, if any, change in the bursting strength, and hence it may be inferred that the error in question is negligible for the present purpose.

The cracks were formed either with a glass-cutter's diamond, or by scratching with a hard steel edge and tapping gently. The subsequent annealing was performed by heating to 450° C. in a resistance furnace, maintaining that temperature for about one hour, and then allowing the whole to cool slowly. The question of the best annealing temperature required careful consideration, as it was evidently necessary to relieve the initial stresses due to cracking as much as possible, while at the same time keeping the temperature so low that appreciable deformation of the crack did not occur. It was found that the bursting strength increased with the annealing temperature up to about 400° C., while between 400° C. and 500° C. very little further change was perceptible. From this it was inferred that relief of the initial stresses was sufficient for the purpose in view at a temperature of 450° C.

---

* By extrapolation.

The principal stresses at rupture, Q and R, were calculated from the observed bursting pressure by means of the usual expressions for the stresses in thin hollow spheres and circular cylinders, the thickness of the glass near the crack being measured after bursting. In the case of the tubes the cracks were parallel to the generators, and provision was made for varying Q by the application of end loads.

In dealing with the longer cracks, leakage was prevented by covering the crack on the inside with celluloid jelly, the tube being burst before the jelly hardened. In the case of the smaller cracks leakage was imperceptible and this precaution was unnecessary.

The time of loading to rupture varied from 30 seconds to five minutes. No evidence was observed of any variation of bursting pressure with time of loading.

The results of the bursting tests are set down in Tables II. and III. below. $2c$ is the length of the crack, Q and R are the calculated principal stresses respectively parallel and perpendicular to the crack, and D is the diameter of the bulb or tube. The thickness of the bulbs was about 0·01 inch and the tubes 0·02 inch.

TABLE II.—Bursting Strength of Cracked Spherical Bulbs.

| $2c$ | D | Q | R | $R\sqrt{c}$. |
|---|---|---|---|---|
| inch. | inch. | lbs. per sq. inch. | lbs. per sq. inch. | |
| 0·15 | 1·49 | 864 | 864 | 237 |
| 0·27 | 1·53 | 623 | 623 | 228 |
| 0·54 | 1·60 | 482 | 482 | 251 |
| 0·89 | 2·00 | 366 | 366 | 244 |

TABLE III.—Bursting Strength of Cracked Circular Tubes.

| $2c$ | D | Q | R | $R\sqrt{c}$. |
|---|---|---|---|---|
| inch. | inch. | lbs. per sq. inch. | lbs. per sq. inch. | |
| 0·25 | 0·59 | −621 | 678 | 240 |
| 0·32 | 0·71 | −176 | 590 | 232 |
| 0·38 | 0·74 | − 31 | 526 | 229 |
| 0·28 | 0·61 | 55 | 655 | 245 |
| 0·26 | 0·62 | 202 | 674 | 243 |
| 0·30 | 0·61 | 308 | 616 | 238 |

The average value of $R\sqrt{c}$ is 239, and the maximum 251.

According to the theory, fracture should not depend on Q, and $R\sqrt{c}$ should have, at fracture, the constant value

$$\sqrt{\frac{2ET}{\pi\sigma}}.$$

In the case of the glass used for these experiments, $E = 9 \cdot 01 \times 10^6$ lbs. per sq. inch, $T = 0 \cdot 0031$ lbs. per inch, and $\sigma = 0 \cdot 251$, so that the above quantity is equal to 266.

These conclusions are sufficiently well borne out by the experimental results, save that the maximum recorded value of $R\sqrt{c}$ is 6 per cent., and the average 10 per cent., below the theoretical value. It must be regarded as improbable that the error in the estimated surface tension is large enough to account for this difference, as this view would render necessary a somewhat unlikely deviation from the linear law.

A more probable explanation is to be obtained from an estimate of the maximum stress in the cracks. An upper limit to the magnitude of the radius of curvature at the ends of the cracks was obtained by inspection of the interference colours shown there. Near the ends a faint brownish tint was observed, and this gradually died out, as the end was approached, until finally nothing at all was visible. It was inferred that the width of the cracks at the ends was not greater than one-quarter of the shortest wave length of visible light, or about $4 \times 10^{-6}$ inch. Hence $\rho$ could not be greater than $2 \times 10^{-6}$ inch.

Taking as an example the last bulb in Table II. and substituting in formula (21), it is found that

$$\frac{a_0}{b_0} = \sqrt{\frac{c}{\rho}} \geq 478$$

$$\frac{2R}{E} = 8 \cdot 13 \times 10^{-5},$$

whence the maximum stress $F \geq 344{,}000$ lbs. per sq. inch. The value given by the first order expression

$$F = 2R\sqrt{\frac{c}{\rho}}$$

is 350,000 lbs. per sq. inch.

A possible explanation of the discrepancy between theory and experiment is now evident. In the tension tests, the verification of HOOKE'S law could only be carried to the breaking stress, 24,900 lbs. per sq. inch. There is no evidence whatever that the law is still applicable at stresses more than ten times as great. It is much more probable that there is a marked reduction in modulus at such stresses. But a decrease in modulus at any point of a body deformed by given surface tractions involves an increase in strain energy, and therefore in the foregoing experiments a decrease in strength. This is in agreement with the observations.

### 5. *Deductions from the Foregoing Results.*

The estimate of maximum stress obtained above appears to lead to a peremptory disproof of the hypothesis that the maximum tension in this glass, at rupture, is always equal to the breaking stress in ordinary tensile tests. If HOOKE's law was obeyed to rupture, and the squares of the strains were negligible, the maximum tension in the above cracked tube could not have been less than 344,000 lbs. per sq. inch; but, in the tensile tests, HOOKE's law was obeyed up to the breaking stress, the squares of the strains were negligible, and the maximum stress was only 24,900 lbs. per sq. inch. Hence the stresses could not have been the same in the two cases. Moreover, the order of the results obtained suggests (though this is not rigorously proved, as the assumptions have not been checked at stresses above 24,900 lbs. per sq. inch) that the actual strength may be more than ten times that given by the hypothesis.

Similar conclusions may be drawn regarding the "maximum extension," "maximum stress-difference" and "maximum shear strain" hypotheses which have been proposed from time to time for estimating the strength of brittle solids.

These conclusions suggest inquiries of the greatest interest. If the strength of this glass, as ordinarily interpreted, is not constant, on what does it depend? What is the greatest possible strength, and can this strength be made available for technical purposes by appropriate treatment of the material? Further, is the strength of other materials governed by similar considerations?

Some indication of the probable maximum strength of this glass may be obtained from the bursting tests already described. There is no reason for supposing that, in those tests, the radii of curvature at the corners of the cracks were as great as $2\times10^{-6}$ inch. It is much more likely that they were of the same order as the molecular dimensions. Considering, as before, the last bulb in Table II., and putting $\rho = 2\times10^{-8}$ inch in formula (21), it is found that the maximum stress, F, is about $3\times10^{6}$ lbs. per sq. inch. Elastic theory cannot, of course, be expected to apply with much accuracy to cases where the dimensions are molecular, on account of the replacement of summation by integration, and the probable diminution of modulus at very high stresses must involve a further error. Taking these circumstances into consideration, however, it may still be said that the probable maximum strength of the glass used in the foregoing experiments is of the order $10^{6}$ lbs. per sq. inch.

It is of interest to enquire at this stage whether there is any reason for ascribing similar maximum strengths to other materials. On the molecular theory of matter the tensile strength of an isotropic solid or liquid is of the same order as, though less than, its "intrinsic pressure," and may therefore be estimated either from a knowledge of the total heat required to vaporise the substance or by means of VAN DER WAAL'S equation.* It may be noted that these methods of estimating the stress indicate that it should be, approximately at least, a constant of the material. TRAUBE* gives the following as the intrinsic pressures of a number of metals, at ordinary temperatures :—

---

* POYNTING and THOMSON, 'Properties of Matter,' ch. xv.

TABLE IV.—Intrinsic Pressures of Metals (TRAUBE).

| Metal. | Intrinsic Pressure. |
|---|---|
| | lbs. per sq. inch. |
| Nickel | $4 \cdot 71 \times 10^6$ |
| Iron | $4 \cdot 70 \times 10^6$ |
| Copper | $3 \cdot 42 \times 10^6$ |
| Silver | $2 \cdot 34 \times 10^6$ |
| Antimony | $1 \cdot 74 \times 10^6$ |
| Zinc | $1 \cdot 58 \times 10^6$ |
| Tin | $1 \cdot 06 \times 10^6$ |
| Lead | $0 \cdot 75 \times 10^6$ |

These are of the same order as the direct estimate obtained above for glass, but they are from 20 to 100 times the strengths found in ordinary tensile and other mechanical tests.

In the case of liquids, the discrepancy between intrinsic pressure and observed tensile strength is much greater. According to VAN DER WAAL'S equation, water has an intrinsic pressure of about 160,000 lbs. per sq. inch, whereas its tensile strength is found to be about 70 lbs. per sq. inch. It has been suggested that this divergence may be due to impurities, such as dissolved air, but DIXON and JOLY† have shown that dissolved air has no measurable effect on the tensile strength of water.

Thus the matters under discussion appear to be of general incidence, in that the strengths usually observed are but a small fraction of the strengths indicated by the molecular theory.

Some further discrepancies between theory and experiment may now be noticed. In the theory it is assumed that rupture occurs in a tensile test at the stress corresponding with the maximum resultant pull which can be exerted between the molecules of the material. On this basis the applied stress must have a maximum value at rupture, and hence, if intermolecular force is a continuous function of molecular spacing, the stress-strain diagram must have zero slope at that point. This, of course, is never observed in tensile tests of brittle materials ; in no case has any evidence been obtained of the existence of such a maximum anywhere near the breaking stress.

Again, the observed differences in strength as between static and alternating stress tests are at first sight inexplicable from the standpoint of the molecular theory, if the breaking load is regarded as the sum of the intermolecular attractions. According to the theory, large changes in the latter can only occur as a result of large changes in the thermal energy of the substance, such as would be immediately evident in alternating stress tests, if they took place.

---

* 'Zeitschr. für Anorganische Chemie,' 1903, vol. xxxiv., p. 413.

† 'Phil. Trans.,' B, 1895, p. 568.

Lastly, as indicated above, the strain energy at rupture of an elastic solid or liquid should on the molecular theory be of the same order as its heat of vaporisation. Hence rupture should be accompanied by phenomena, such as a large rise of temperature, indicative of the dissipation of an amount of energy of this order. It is well known that tensile tests of brittle materials show no such phenomena.

If, as is usually supposed, the materials concerned are substantially isotropic, there is but one hypothesis which is capable of reconciling all these apparently contradictory results. The theoretical deduction—that rupture of an isotropic elastic material always occurs at a certain maximum tension—is doubtless correct; but in ordinary tensile and other tests designed to secure uniform stress, the stress is actually far from uniform so that the average stress at rupture is much below the true strength of the material.

Now it may be shown, with the help of elastic theory, that the stress must be substantially uniform, in such tests, unless the material of the test-pieces is heterogeneous or discontinuous. It is known that all substances are in fact discontinuous, in that they are composed of molecules of finite size, and it may be asked whether this type of discontinuity is sufficient to account for the observed phenomena.

With the help of formula (13) above, this question may be answered in the negative. Formula (13) shows that a thin plate of glass, having in it the weakest possible crack of length $2c$ inch, will break at a tension, normal to the crack, of not less than $266/\sqrt{c}$ lbs. per sq. inch. This result, however, is subject to certain errors, and experiment shows that the true breaking stress is about $240/\sqrt{c}$ lbs. per sq. inch. But such a crack is the most extreme type, either of discontinuity or heterogeneity, which can exist in the material. Hence it is impossible to account for the observed strength, 24,900 lbs. per sq. inch, of the simple tension test specimens, unless they contain discontinuities at least $2 \times \left(\frac{240}{24,900}\right)^2$ inch, or say, $2 \times 10^{-4}$ inch wide. This is of the order $10^{-4}$ times the molecular spacing.

The general conclusion may be drawn that the weakness of isotropic solids, as ordinarily met with, is due to the presence of discontinuities, or flaws, as they may be more correctly called, whose ruling dimensions are large compared with molecular distances. The effective strength of technical materials might be increased 10 or 20 times at least if these flaws could be eliminated.

It is easy to see why the presence of such small flaws can leave the strength of cracked plates, such as those of the foregoing experiments, practically unaffected. The most extreme case of weakening is that where there is a flaw very near the end of the crack and collinear with it. Here the result is merely to increase the effective length of the crack by less than $10^{-8}$ inch. This involves a weakening of less than 0·1 per cent.

### 6. *The Strength of Thin Fibres.*

Consideration of the consequences of the foregoing general deduction indicated that very small solids of given form, *e.g.*, wires or fibres, might be expected to be stronger than large ones, as there must in such cases be some additional restriction on the size of the flaws. In the limit, in fact, a fibre consisting of a single line of molecules must possess the theoretical molecular tensile strength. In this connection it is, of course, well known that fine wires are stronger than thick ones, but the present view suggests that in sufficiently fine wires the effect should be enormously greater than is observed in ordinary cases.

This conclusion has been verified experimentally for the glass used in the previous tests, strengths of the same order as the theoretical tenacity having been observed. Incidentally, information of interest has been obtained, somewhat unexpectedly, concerning the genesis of the flaws, and it has been found to be possible to prepare quite thick fibres in an unstable condition in which they have the theoretical strength.

Fibres of glass, about 2 inches long and of various diameters, were prepared. One end of a fibre was attached to a stout wire hanging on one arm of a balance, and the other end to a fixed point, the medium of attachment being sealing wax. A slight tension was applied while the wax was still soft, in order to eliminate bending of the fibre at the points of attachment. The other arm of the balance carried a beaker into which water was introduced from a pipette or burette. The weight of water necessary to break the fibre was observed, and the diameter of the latter at the fracture was found by means of a high-power measuring microscope. Hence the tensile strength was obtained.

At first the results were extremely irregular, though the general tendency of the strength to increase with diminishing diameter was clear. It was found that the irregularities were due to the dependence of the strength on the following factors :—

(1) The maximum temperature of the glass.—To secure the best results it was found necessary to heat the glass bead to about 1400° C. to 1500° C. before drawing the fibre.

(2) The temperature during drawing.—If the glass became too cool before drawing was complete, a weak fibre was obtained. This temperature could not be very closely defined, but it is perhaps the same as the limiting temperature of the viscous liquid phase, namely, 730° C. This effect made the drawing of very fine fibres a matter of some difficulty, as the cooling was so rapid.

(3) The presence of impurities and foreign bodies.

(4) The age of the fibre.—For a few seconds after preparation, the strength of a properly treated fibre, whatever its diameter, was found to be extremely high. Tensile strengths ranging from 220,000 to 900,000 lbs. per sq. inch were observed in fibres up to about 0·02 inch diameter. These strengths were estimated by measuring the radii to which it was necessary to bend the fibres in order to break them. They are therefore probably somewhat higher than the actual tenacities. The glass appeared to be

almost perfectly elastic up to these high stresses. The strength diminished, however, as time went on, until after the lapse of a few hours it reached a steady value whose magnitude depended on the diameter of the fibre.

Similar phenomena have been observed with other kinds of glass, and also with fused silica.

The relation between diameter and strength in the steady state was investigated in the following manner. Fibres of diameters ranging from 0·13 × $10^{-3}$ inch to 4·2 × $10^{-3}$ inch, and 6 inches long, were prepared by heating the glass to about 1400° C. to 1500° C. in an oxygen and coal-gas flame and drawing the fibre by hand as quickly as possible. The fibres were then put aside for about 40 hours, so that they might reach the steady state. The test specimens were prepared by breaking these fibres in tension several times until pieces about 0·5-inch long remained; these were then tested by the balance method already described. The object of this procedure was the elimination of weak places due to minute foreign bodies, local impurities and other causes.

Table V. below gives the results of these tests. Diameters are in thousandths of an inch, and breaking stresses in lbs. per sq. inch.

Table V.—Strength of Glass Fibres.

| Diameter. | Breaking Stress. | Diameter. | Breaking Stress. |
|---|---|---|---|
| 0·001 inch. | lbs. per sq. inch. | 0·001 inch. | lbs. per sq. inch. |
| 40·00 | 24,900* | 0·95 | 117,000 |
| 4·20 | 42,300 | 0·75 | 134,000 |
| 2·78 | 50,800 | 0·70 | 164,000 |
| 2·25 | 64,100 | 0·60 | 185,000 |
| 2·00 | 79,600 | 0·56 | 154,000 |
| 1·85 | 88,500 | 0·50 | 195,000 |
| 1·75 | 82,600 | 0·38 | 232,000 |
| 1·40 | 85,200 | 0·26 | 332,000 |
| 1·32 | 99,500 | 0·165 | 498,000 |
| 1·15 | 88,700 | 0·130 | 491,000 |

It will be seen that the results are still somewhat irregular. No doubt more precise treatment of the fibres would lead to some improvement in this respect, but such refinement is scarcely necessary at the present stage.

The limiting tensile strength of a fibre of the smallest possible (molecular) diameter may be obtained approximately from the figures in Table V. by plotting reciprocals of the tensile strength and extrapolating to zero diameter. This maximum strength is found to be about 1·6 × $10^6$ lbs. per sq. inch, which agrees sufficiently well with the rough estimate previously obtained from the cracked plate experiments.

---

* From the tensile tests previously described.

In 1858, KARMARSCH* found that the tensile strength of metal wires could be represented within a few per cent. by an expression of the type

$$F = A + \frac{B}{d} \quad \text{. . . . . . . . . . . . .} \quad (22)$$

where $d$ is the diameter and A and B are constants. In this connection it is of interest to notice that the figures in Table V. are given within the limits of experimental error by the formula

$$F = 22{,}400 \frac{4 \, . \, 4 + d}{0 \, . \, 06 + d} \quad \text{. . . . . . . . . . .} \quad (23)$$

where F is in lbs. per sq. inch and $d$ is in thousandths of an inch. Within the range of diameters available to KARMARSCH, this expression differs little from

$$F = 22{,}400 + \frac{98{,}600}{d}, \quad \text{. . . . . . . . . . .} \quad (24)$$

which is of the form given by KARMARSCH. Moreover, the values of B found by him for the weaker metals, *e.g.*, silver and gold, in the annealed state, are of the same order as that given by formula (24) for glass.

To a certain extent this correspondence suggests that the mechanism of rupture, as distinct from plastic flow, in metals, is essentially similar to that in brittle amorphous solids such as glass.

The remarkable properties of the unstable strong state referred to on p. 180 above are exhibited most readily in the case of clear fused silica. If a portion of a rod about 5 mm. diameter be made as hot as possible in an oxyhydrogen flame, and then drawn down to, say, 1 mm. or less and allowed to cool, the drawn-down portion may be bent to a radius of 4 or 5 mm. without breaking, and if then released will spring back almost exactly to its initial form. If instead of being released it is held in the bent form it will break spontaneously after a time which usually varies from a few seconds to a few minutes, according to the degree of flexure. To secure the best results the drawing should be performed somewhat slowly.

When fracture occurs it is accompanied by phenomena altogether different from those associated with the fracture of the normal substance. The report is much louder and deeper than the sharp crack which accompanies the rupture of an ordinary silica or glass rod, and the specimen is invariably shattered into a number of pieces, parts being frequently reduced to powder. This shattering is not confined to the highly stressed drawn-out fibre; it usually occurs also in the unchanged parts of the original thick rod and sometimes in the free ends, which are not subjected to the bending moment. The experiment is most striking, for it appears at first sight that the 5 mm. rod has been broken by a couple applied through the fibre, which may be only 0·5 mm. in diameter. As a matter of fact, however, the shattering is probably merely one of the means of dissipating the strain energy of the strong fibre, which at fracture is perhaps 10,000 times that of silica in the ordinary weak state. An elastic wave is doubtless propagated from the original fracture, and the stresses due to this wave shatter the rod.

---

* 'Mittheilungen des gew. Ver. für Hannover,' 1858, pp. 138–155.

Confirmation of this view is obtainable if the fibre is broken by twisting instead of by bending. The thick part of the rod is in this case found to contain a number of spiral cracks, at an angle of about 45° to the axis, showing that the material has broken in tension, but the cracks run in both right- and left-handed spirals, so that the surface of the rod is divided up into little squares. This shows that the cracking must be due to an alternating stress, such as would result from the propagation of a torsional wave along the rod.

Another phenomenon which has been observed in these fibres is that fracture at any point appears to cause a sudden large reduction in the strength of the remaining pieces. Thus, in one case a glass fibre was found to break in bending at an estimated stress of 220,000 lbs. per sq. inch. One of the pieces, on being tested immediately afterwards, broke at about 67,000 lbs. per sq. inch.

### 7. *Molecular Theory of Strength Phenomena.*

From the engineering standpoint the chief interest of the foregoing work centres round the suggestion that enormous improvement is possible in the properties of structural materials. Of secondary, but still considerable, importance is the demonstration that the methods of strength estimation in common use may lead in some cases to serious error.

Questions relating to methods of securing the indicated increase in tenacity, or of eliminating the uncertainty in strength calculations, can scarcely be answered without some more or less definite knowledge of the way in which the properties of molecules enter into the phenomena under consideration. In this connection it is of interest to enquire whether any indication can be obtained of the nature of the properties which are requisite for an explanation of the observed facts.

For this purpose it is convenient to start with molecules of the classical type, whose properties may be defined as (*a*) a central attraction between each pair of molecules which decreases rapidly as their central distance increases, and which depends only on that distance and the nature of the molecules ; (*b*) translational and possibly rotational vibrations whose energy is the thermal energy of the substance. In the unstrained state, the kinetic reactions due to (*b*) balance the central attractions (*a*).

In a body composed of such molecules, the flaws which have been shown to exist in real substances might consist of actual cracks. But experiment shows that under certain conditions the strength of glass diminishes with lapse of time. On the present hypothesis this would require the potential energy of the system to increase spontaneously by the amount of the surface energy of the cracks. This view must therefore be regarded as untenable.

Again, the observed weakening might conceivably occur if at any instant the vibrations of a large number (at least $10^8$) of near molecules synchronised and were in phase, provided the energy of these molecules was approximately that corresponding with the temperature of ebullition of the substance. Except in the case of a material very near its boiling point, the probability of such an occurrence must be so small as to be quite negligible. Hence this hypothesis also must be discarded.

The foregoing discussion seems to suggest that the assumed type of molecule is too simple to permit of the construction of an adequate theory. An increase in generality may be obtained by supposing that the attraction between a pair of molecules depends not only on their distance apart, but also on their relative orientation. The properties of crystals seem definitely to require the molecules of anisotropic materials to be of this type, but those of isotropic substances have usually been assumed to be of the simpler kind. In view of the author, however, molecular attraction must be a function of orientation even in substances, such as glass, metals and water, which are usually referred to as "isotropic."

Consider a solid made up of a number of such molecules, initially oriented at random. Doubtless the mechanical properties of the substance, while it is in this amorphous condition, will differ little from those of a substance composed of molecules of the simpler type, having an attraction of appropriate strength. If this is so, the tensile strength of the material must be that corresponding with its average intrinsic pressure.

In general, however, this initial condition cannot be one of minimum potential energy.

It is clear that under suitable conditions the tendency to attain stable equilibrium can cause the molecules to rotate and set themselves in chains or sheets, with their maxima of attraction in line. The formation of sheets will commence at a great number of places throughout the solid, *i.e.*, wherever the initial random arrangement is sufficiently favourable. Evidently it is possible for the number of such "sufficiently favourable" arrangements to be enormously less than the total number of molecules, so that the ultimate result will be the formation of a number of units or groups, each containing a large number of molecules oriented according to some definite law. The relative arrangement of the units will, of course, be haphazard.

Now, in each unit there will, in general, be a direction which is, approximately at least, that of the minimum attractions of the majority of the molecules in the unit. Hence if the ratio of the maximum to the minimum attractions is sufficiently great, each unit can constitute a "flaw," and there appears to be no reason why the units should not be as large as the flaws have been shown to be in the case of glass. Thus, in order to explain the spontaneous weakening of glass, it is only necessary to suppose that the thermal agitation at about 1400° C. is sufficient to bring about the initial random formation.

It will be remembered that in the case of the freshly drawn fibres the reduction in tenacity required several hours for completion, so that the time taken was large compared with the time of cooling. Expressed in terms of molecular motion, this means that the molecules resist rotation very much more than they resist translation. This is in keeping with the conclusions of DEBYE,* who found that, on the basis of the quantum theory, the phenomena associated with the specific heat of solids could be explained only if the thermal vibrations of the molecules were regarded as practically irrotational. The same thing is shown more roughly, without introducing the quantum theory, by the law of DULONG and PETIT, which requires that each molecule shall have only three degrees of freedom.

The theory here put forward makes the spontaneous weakening a consequence of the attainment of a molecular configuration of stable equilibrium ; it therefore suggests that the weakening should be accompanied, in general, by a change in the dimensions of the solid. This has been verified by direct observation with a high-power microscope ; in the course of half an hour a spontaneously weakening glass fibre increased in length by about 0·1 per cent., while the length of a silica fibre decreased by about 0·03 per cent.

On account of the random arrangement of the molecular groups, this spontaneous change in unstrained volume must set up internal stresses, which may be sufficiently large to start cracks along the directions of least strength. In this connection it may be mentioned that irregularly shaped pieces of glass, of which some parts had been put into the strong unstable state by heating, have sometimes been observed to break spontaneously about an hour after cooling was practically complete.

It was remarked on p. 184 that cracks could not form spontaneously in a substance composed of molecules having spherical fields of force, as the process would involve an increase in potential energy. This is no longer true when the attraction is a function of orientation, as the surface energy of the cracks may be more than counterbalanced by the decrease in potential energy accompanying the molecular rearrangement.

For this reason, it is impossible to deduce the ratio of the maximum to the minimum molecular attractions from the ratio of the maximum and minimum strengths of the material, as it is possible that the spontaneous weakening is always accompanied by the formation of minute cracks, of the same size as the molecular groups.

It is probable that, in many cases, the most stable orientation of the molecules at a free surface is that in which their maxima of attraction lie along the surface. Such an orientation would in turn lead to a similar tendency on the part of the next layer of molecules, and so on, the tendency diminishing with increasing distance from the surface. There would therefore be a surface layer having the special property that in it the " flaws " ran parallel to the surface.

Hence this layer would be of exceptional strength in the direction of the surface. This suggests a reason for the experimental fact that the breaking load of wires and fibres consists mainly of two parts, one proportional to the area, and the other to the perimeter of the cross-section. The process of drawing, too, might predispose the molecules to take up positions with their maxima of attraction parallel to the surface.

---

* 'Ann. der Physik,' 39 (1912), p. 789.

If a perfectly clean glass plate be covered with gelatine and set aside, the gelatine gradually contracts, and as it does so it tears from the glass surface thin flakes up to about 0·06-inch diameter and shaped like oyster shells.* This tendency to flake at the surface is also observed when glass is broken by bending. This was particularly well shown in the specially prepared fibres used for the experiments described in the present paper. In almost all cases of flexural fracture the crack curled round on approaching the compression side, till it was nearly parallel to the surface. On two occasions the fracture divided before changing direction, the two branches going opposite ways along the fibre and a flake of length several times the diameter of the fibre was detached.

Surface flaking is also observed when some kinds of steel are subjected to repeated stress. Here the flakes are usually very small.

All these facts are evidently in complete agreement with the "surface layer" theory and, indeed, it is difficult to account for them on any other basis.

### 8. *Extended Application of the Molecular Orientation Theory.*

On the basis of the present theory, the physical properties of materials must be intimately related to the geometrical properties of the molecular sheet-formation. In order that a substance may exhibit the characteristic properties of crystals, it is clearly necessary for the sheets of molecules to be plane. In this case the crystals are, of course, the molecule groups or "units" referred to above. In "amorphous" materials, on the other hand, the sheets are probably curved.†

In materials of the former type, there must exist planes on which, if they are subjected to a sufficiently large shearing stress, the portions on either side of the planes can undergo a mutual sliding through a distance equal to any integral multiple of the molecular spacing, without fundamentally affecting the structure of the crystal. It is well known that the phenomenon of yield in crystals, and especially in metals, is of this nature. The planes in question are, of course, the well-known "gliding planes," and it is further possible that they may be identified also with the surfaces of least attraction. The stress at which gliding occurs in a single crystal must be determined in the following manner. The molecules of a crystal are normally in a configuration of stable equilibrium, and if two parts of the crystal slide on a gliding plane through one molecular space the resulting configuration is also stable. Between these two positions there must, in general, be one of higher potential energy, in which the equilibrium is unstable, and the shearing stress is determined by the condition that the rate at which work is done, in sliding from the stable to the unstable state, must be equal to the greatest rate of increase in potential energy which occurs during the passage between the two states. This rate will depend on the shape of the molecular fields of force, and may in particular cases be zero. Liquid crystals are doubtless of this type. The average shear stress, during yield, of a random aggregation of a large number of crystals, is doubtless greater than that of a single crystal, as the angle between the gliding planes and the maximum shear stress must vary from crystal to crystal and can be zero in only a few of them.

---

* Lord Rayleigh, 'Engineering,' 1917, vol. 103, p. 111, and H. E. Head, 'Engineering,' 1917, vol. 103, p. 138.

† See Quincke, 'Ann. der Physik,' (4), 46, 1915, p. 1025.

As the mutually gliding portions of a crystal pass from the stable to the unstable state, the molecular cohesion between them (normal to the gliding plane) must, in general, become less. In particular instances it may diminish to zero before the position of unstable equilibrium is reached. In these cases, shearing fracture along the gliding planes will occur, unless the material is subjected to a sufficiently high "hydrostatic" pressure, in addition to the shearing stress. Thus, a crystalline substance may be either ductile or brittle, according to nature of the applied stress, or it may be ductile at some temperatures and brittle at others, under the same kind of stress as has been actually observed by BENGOUGH and HANSON in the case of tensile tests of copper. This rupture in shear explains the characteristic fracture of short columns of brittle crystalline material under axial compression. The theory indicates that such fracture can always be prevented and yield set up by applying sufficient lateral pressure in addition to the longitudinal load; this is in agreement with experiments on rocks such as marble and sandstone.* Conversely, a ductile substance might be made brittle if it were possible to apply to it a sufficiently large hydrostatic tension.

In the case of an alloy of, say, two metals A and B, suppose, as an example, that the sequence of molecules on either side of a gliding plane is

. A . B . B . A . B . B
B . B . A . B . B . A .

Let sliding occur (through one molecular space) to an adjoining position of stable equilibrium, or, say, to the configuration

. A . B . B . A . B . B . ←
. B . B . A . B . B . A →

Evidently, the structure in the neighbourhood of the gliding plane is in this case no longer the same as in the original crystal formation. It is therefore likely that the new state is one of higher potential energy, whence it is reasonable to suppose that the maximum rate of increase in potential, in sliding, is greater than it would have been had the potential of the two states been the same. Thus an alloy may be expected to have a higher yield-point than its most ductile constituent. This is in accordance with experience. For example, it is known that quenching from a high temperature hardens tool steel by preventing the separation of "ferrite," or iron containing no carbon.

---

* T. V. KARMAN, 'Zeitschr. Ver. Deutsch. Ing.,' 55, 1911.

In a single crystal the molecules are presumably in an equilibrium configuration of maximum stability. In this event, the equilibrium of molecules at or near intercrystalline boundaries, in a body composed of a large number of crystals, must, in general, be less stable than that of the molecules in the interior of the crystals. In fact, where the orientation of the component crystals is haphazard, the stability of the boundary molecules may be expected to range from the maximum of normal crystallisation down to zero, *i.e.*, neutral equilibrium. If such a body be subjected to a shear stress, some of the molecules in or near neutral equilibrium must, in general, become unstable, and these will tend to rotate to new positions of equilibrium. This rotation, however, will be strongly resisted, as has been seen, by forces doubtless of a viscous nature, and its amount will accordingly depend on the time during which the stress is applied. If, therefore, the strain is observed it will be found to increase slowly as time goes on, but at a constantly decreasing rate, as the molecules concerned approach equilibrium. If now the load is removed, these molecules must rotate in order to regain their original positions of equilibrium, and this process in turn will be retarded by viscous forces. Hence a small part of the observed strain will remain after the removal of the load, and this will gradually disappear as time goes on. These properties, known as "elastic after-working," are, of course, well known to belong to crystalline materials. Moreover, the theory shows that they should not be possessed by single crystals, and this has been demonstrated experimentally.*

There is a special type of gliding or yield which may occur at stresses below the normal yield point. Consider a pair of adjacent crystals, separated by a plane boundary. If these crystals are thought of as sliding relatively to each other, it will be seen that only in a finite number of the positions so taken up can the two be in stable equilibrium. Between each pair of such positions there must in general be one of unstable equilibrium. Suppose that, while near such an unstable position, the two crystals are embedded in a number of others. Under these conditions the boundary molecules of the two crystals will be pulled over in the direction of one or other of the two adjoining stable positions, and they will strain the solid in the process. If now the body is subjected to a shearing stress tending to cause relative displacement of the two crystals towards the other stable position, then at a certain value of this stress the molecules on either side of the boundary will be wrenched away, will pass through the position of instability, and will then take up a new position bearing the same relation to the second stable position as their original state did to the first. This new condition will, of course, persist after the removal of the load, as the original state cannot be regained without passing through unstable equilibrium, *i.e.*, a condition of maximum potential energy. To cause the crystals to pass through this condition it would be necessary to apply a load of opposite sign, and in this way the process might be repeated indefinitely. In a body composed of a large number of crystals there must be many arrangements of this type, in which adjacent crystals can execute inelastic oscillations about positions of unstable equilibrium, under alternating shear stresses below the ordinary yield stress. The consequent observed phenomena would correspond exactly with those known to be manifested in metals, under the name "elastic hysteresis."*

---

* H. v. WARTENBERG, 'Deutsch. Phys. Gesell., Verh.,' 20, pp. 113–122, August 30, 1918.

Experimentally, elastic hysteresis is distinguished from elastic after-working by the circumstance that it is completed very much more quickly. This is just what would be expected theoretically, on the view that molecular translation occurs much more readily than rotation.

It has been remarked that when a single crystal of a pure substance is caused to yield, its structure is fundamentally unaltered. This cannot hold, however, in the case of an aggregate of a large number of crystals arranged at random, or a crystal embedded in amorphous material. True, the material in the interior of each crystal can retain its original properties, but near the crystalline boundaries the structure must be violently distorted. As a result, it may be expected that the number of the molecules of inferior stability will be largely increased. Elastic after-working in metals should therefore be increased by overstraining or "cold-working." This, again, agrees with experience.

The foregoing considerations lend support to the view that each crystal of a severely cold-worked piece of metal is surrounded by an amorphous layer of appreciable thickness. If such a piece of metal undergoes a shear strain greater than that which can initiate yield in the normal crystalline substance, the average stress which is set up must be above the normal yield stress, for the part due to the amorphous layers must be the elastic stress corresponding with the strain, and this, by hypothesis, is greater than the yield stress. This part, moreover, will increase with the strain. It follows that yield in cold-worked metal should be less sharply defined, and should occur at a higher shear stress than in the normal crystalline variety. That this is actually the case is, of course, well known.

In the case of very large strains an important part of the shear stress must be taken by the amorphous boundary layers, and as a result the maximum tensile stress may reach a value sufficient to cause rupture of some favourably disposed crystals across their planes of least strength. This is, perhaps, the actual mode of rupture in ductile materials. On this view, the "ductility" of a metal depends simply on the relation between the tensile strength of the "flaws" and the normal yield stress. A substance whose ductility is small may still be "malleable," as hammering need not give rise to large tensile stresses.

The formation of non-crystalline material at the intercrystalline boundaries, when a piece of metal is over-strained, appears to provide an explanation of the sudden drop in stress which occurs immediately after the initiation of yield† in ductile metals.

---

* **Guest** and **Lea**, "Torsional Hysteresis of Mild Steel," 'Roy. Soc. Proc.,' A, June, 1917.

† **Robertson** and **Cook**, 'Roy. Soc. Proc.,' A, vol. 88, 1913, pp. 462–471.

Remembering that the surface tension of a substance is the work done in forming unit area of new surface, it will be seen that the tension of any surface of a crystal must depend on the angle it makes with the crystal axis. Thus the surface tension parallel to the planes of least strength must be less than that in any other direction. Consequently, in a body composed of a number of crystals there must exist a mutual surface tension at each intercrystalline boundary. Now, the theory of surface tension shows that the magnitude of such a mutual tension is greatly diminished by making the transition between the two bodies more gradual. Hence the formation of the amorphous boundary layer involves a reduction in the surface energy of the crystals, and this is shown in the experiments by a drop in the stress. If this account of the phenomenon is complete, the drop in stress must be determined by the condition that the loss of strain energy equals the reduction in surface energy. The mechanism of the process appears to be that the breaking up of the boundary, which must accompany yield, is resisted by the surface tension, and yielding therefore requires a higher stress for its initiation than for its maintenance.

According to this view, the loss of strain energy should be inversely proportional to the linear dimensions of the crystals. Hence the results of different experiments should show considerable variation in the magnitude of the drop in stress. This is actually the case ; a single series of experiments on mild steel, by ROBERTSON and COOK, gave drops varying from 17 per cent. to 36 per cent., while in other experiments as little as 7 per cent. has been observed.

In the above series of experiments the average loss of strain energy was about 12 inch-lbs. per cubic inch. Assuming, for simplicity, that the crystals were cubes, of, say, 0·001-inch side (which is a fair value for well-treated mild steel), the area of the intercrystal surface was 3000 sq. inches per cubic inch. These figures give the average intercrystal surface tension as 0·004 lbs. per inch. This is certainly of the right order of magnitude.

Many of the phenomena discussed above will be more complicated, in practice, if the coefficient of expansion of the crystals is not the same in all directions. In such an event, internal stresses will be set up in cooling, on account of the random arrangement of the crystals, and these stresses must be taken into consideration in applying the theory.

There remains for consideration the problem of the fracture of metals under alternating stress. It is known that fatigue failure occurs as the result of cracking after repeated slipping on gliding planes, and the theory has been advanced* that this cracking is due to repeated to and fro sliding and consequent attrition and removal of material from the gliding planes. This theory presents some difficulties, in that it does not explain how the attrition can occur, or the method of disposing of the debris.

A theory which is free from these objections may be constructed if it is supposed that a change in volume occurs on the passage of the metal from the crystalline to the amorphous state. This assumption is, of course, known to be valid for many substances at their melting points, but at lower temperatures there seems to be no definite information available.

---

* EWING and HUMFREY, 'Phil. Trans.,' A, 1902, p. 200.

This assumption being granted, suppose that a piece of material which contracts on decrystallising is being subjected to a stress cycle just sufficient to cause repeated slipping in the most favourably disposed crystals. As a result, the material at the boundaries of these crystals will become amorphous, and the quantity of amorphous material will increase continuously as long as the repeated slipping goes on. But, by hypothesis, the unstrained volume of the amorphous phase is less than the space it filled when in the crystalline state. Hence all the material in the immediate neighbourhood will be subjected to a tensile stress, and as soon as this exceeds a certain critical value a crack will form. It has been observed above that the application of a sufficiently large hydrostatic tension may be expected to make a ductile substance brittle. Hence the crack may occur either in tension or in shear, according to the properties of the material and the nature of the applied stress. Further alternations of stress will cause this crack to spread until complete rupture occurs. This theory makes the limiting safe range of stress equal to that which just fails to maintain repeated sliding in the most favourably disposed crystals.

It may be asked why such cracking does not take place in a static test where the quantity of amorphous material, once yield has fairly started, is presumably much greater. The answer to this is two-fold. In the first place, if the material becomes amorphous round all, or nearly all, the crystals of a piece of metal, it is evident that it will contract as a whole and no great tensile stress will be set up. In the case where only a few crystals yield, the tension arises from the rigidity of the unchanged surrounding metal.

In the second place, even if some crystals do crack, the cracks will not, in general, tend to spread through the ductile cores of the neighbouring crystals, unless the applied load is alternating, on account of the equalisation of stress due to yield.

The safe limit of alternating stress will usually be less than the apparent stress necessary to initiate yield in a static test, on account of initial stresses, including those due to unequal contraction of the crystals.

The theory indicates that the cracking of the first crystal marks a critical point in the history of the piece. At any earlier stage the effects of the previous loading may be removed by heat treatment, or possibly by a rest interval, but once a crack has formed this cannot be done. True, the tension may be relieved and the ruptured crystal may even be compressed somewhat, but this cannot, in general, close the crack, as cracking is not a "reversible" operation. An exception may occur if the top temperature of the heat-treatment is sufficient to bring the molecules on either side of the crack within mutual range by thermal agitation, but it is unlikely that this can happen save in the case of very small cracks.

If this theory is correct, it appears at first sight that the phenomenon of fatigue failure must be confined to substances which contract on decrystallising. This, however is not necessarily so. If, for instance, a small thickness of material at the interface between two crystals were to increase in volume, it could not be said without proof that tensile stresses would not be set up thereby, in addition to compressions. In some cases, in fact, it is obvious that there must be tensions. Thus, if the outer layer of a sphere increases in volume, the matter inside must be subjected to a tensile stress.

The effect of overstrain on the density of metals is at present under investigation at the Royal Aircraft Establishment. The work is not yet sufficiently complete for detailed publication, but it may be mentioned here that the expected change in density has been found, and that the results already obtained are such as to leave little room for doubt that this change is in fact the cause of fatigue failure in metals. Thus, in overstraining mild steel by means of a pure shearing stress, a decrease in average density of as much as 0·25 per cent. has been observed.

Some progress has also been made in the direction of estimating the internal stresses set up as a result of the change in density, and it has been found that an average change of the magnitude mentioned above could give rise to a hydrostatic tensile stress in the cores of the crystals, of the order of 30,000 lbs. per sq. inch.

Dealing now with materials whose molecular sheet-formations are curved, it is at once evident that all yield, or slide, phenomena must be absent, as possible gliding planes do not exist. Thus, this case, though geometrically more complicated, is practically much simpler than that in which the sheets are plane. The theoretical properties of materials having the curved type of formation appear to correspond exactly with those known to belong to brittle "amorphous" substances. Exactly as in the case of crystalline materials, elastic after-working is explained by the inferor stability of molecules near the boundaries of the units of molecular configuration, but elastic hysteresis should not occur. If adequate precautions are taken to avoid secondary tensile stresses, fracture of short columns in compression should occur at stresses of an altogether higher order than in the case of crystals. In this connection it may be remarked that the compressive strength of fused silica is about 25 times as great as its ordinary tensile strength.

It appears from the foregoing discussion that the molecular orientation theory is capable of giving a satisfactory general account of many phenomena relating to the mechanical properties of solids, though closer investigation will perhaps show that the agreement is in some cases superficial only. Such questions as the effects of unequal cooling, foreign inclusions and local impurities, and the behaviour of mixtures of different crystals, have not been dealt with; it is thought that these are matters of detail whose discussion cannot usefully precede the establishment of the general principles on which they depend.

### 9. *Practical Limitations of the Elastic Theory.*

It is now possible to indicate the directions in which the ordinary mathematical theory of elasticity may be expected to fail when applied to real solids.

It is a fundamental assumption of the mathematical theory that it is legitimate to replace summation of the molecular forces by integration. In general this can only be true if the smallest material dimension, involved in the calculations, is large compared with the unit of structure of the substance. In crystalline metals the crystals appear, from the foregoing investigation, to be anisotropic and they must therefore be regarded as the units of structure. Hence the theory of isotropic homogeneous solids may break down if applied to metals in cases where the smallest linear dimension involved is not many times the length of a crystal.

Similar considerations apply to solids such as glass, save that here the units of structure are probably curved.

The most important practical case of failure is that of a re-entrant angle or groove. Here the theory may break down if the radius of curvature of the re-entering corner is but a small number of crystals long. An extreme instance is that of a surface scratch, where the radius of curvature may be but a fraction of the length of the crystals.

In the case of brittle materials the general nature of the effect of scratches on strength may be inferred from the theoretical criterion of rupture enunciated in section 2 above. Whether the material be isotropic or anisotropic, homogeneous or heterogeneous, it is necessary on dimensional grounds that the strain energy shall depend on a higher power of the depth of the scratch than the surface energy. It follows that small scratches must reduce the strength less than large ones of the same shape. Hence, where the tenacity of the material, under "uniform" stress, is determined by the presence of "flaws," it must always be possible to find a certain depth of scratch whose breaking stress is equal to that of the flaws. Evidently such a scratch can have no influence on the strength of the piece. Deeper scratches must have some weakening effect, which must increase with the depth, until in the limit the strength of very large grooves may be found by means of the elastic theory and the appropriate empirical hypothesis of rupture.

In the case of ductile metals, the effect of scratches is important only under alternating or repeated stresses. On the theory advanced in the preceding section, fatigue failure under such stresses is determined by phenomena which occur at the intercrystalline boundaries. Hence the strength of a scratched piece is fixed, not by the maximum stress range in the corner of the scratch, but by the stress range at a certain distance below the surface. This distance cannot be less than the width of one crystal, and it may be greater. Elastic theory suggests that the stress due to a scratch falls off very rapidly with increasing distance from the re-entrant corner, so that the relatively small effect of scratches in fatigue tests is readily explained.

Possibly many published results bearing on this matter depend more on initial skin stresses than on sharp corner effects.

### 10. *Methods of Increasing the Strength of Materials.*

The most obvious means of making the theoretical molecular tenacity available for technical purposes is to break up the molecular sheet-formation and so eliminate the flaws." In the case of crystalline material this has the further advantage of eliminating yield and probably also fatigue failure.

In materials which normally have curved sheets, the molecular fields of force are presumably asymmetrical, and the process indicated above would lead of necessity to a random arrangement, which might be unstable. It has been seen that in glass and fused silica it is actually unstable, except in the case of the finest fibres.

As regards crystalline materials, however, in which the fields of force must have some sort of symmetry, there seems to be no reason why there should not be possible a very fine grained stable configuration, which could be derived from the ordinary crystalline form by appropriate rotations of certain molecules to new positions of stable equilibrium, in such a way as to break up the gliding planes. The grain of such a structure need be but a few molecules long, and its strength would approximate to the theoretical value corresponding with the heat of vaporisation.

There is some evidence that mild steel which has been put into the amorphous condition by over-strain tends, under certain conditions, to take up a stable fine-grained formation of this kind, in preference to resuming its original coarse crystalline configuration, in that a temperature of 0° C. appears to prevent recovery from tensile over-strain.*

These considerations suggest that if a piece of metal were rendered completely amorphous by cold-working, and then suitably heat-treated, its molecules might take up the stable strong configuration already described. The theory indicates, however, that over-straining tends to set up tensile stresses in the unchanged parts of the crystals which may start cracks long before decrystallisation is complete. Such cracking could be prevented if the over-straining were carried out under a sufficiently great hydrostatic pressure, and this line of research seems to be well worth following up. It might, of course, be found that the requisite pressure was so enormous as to render the method unworkable, but if the theory is sound there seems to be no other reason why definite results should not be obtained.

The problem may be attacked in another way. As has been seen, the theory suggests that the drop in stress at the initiation of yield is due to the surface energy of the inter-crystal boundaries. Thus the yield point may be raised by "refining" the metal, *i.e.*, so heat-treating it as to reduce the size of the crystals. The limit of refinement is, doubtless, reached when each "crystal" contains but a single molecule and the material is then in the strong stable state already described.

Refining is also of great value in connection with resistance to fatigue failure. Suppose, in accordance with the foregoing theory of fatigue, that one crystal has been fractured, then the general criterion of rupture shows that the crack cannot spread unless the material is subjected to a certain minimum stress, which is greater the smaller the crack. Thus, reducing the size of the crystals increases the stress necessary to cause the initial crack to extend. There is therefore a critical size of crystal for which the stress-range necessary to spread the crack is equal to that necessary to start it. Until the refining has reached that stage it can have no effect on the magnitude of the safe stress range, but from that point on the range must increase progressively with refinement until the limit is reached, as before, when each "crystal" contains but one molecule.

---

* COKER, 'Phys. Rev.,' 15, August, 1902.

It therefore appears that refining is one avenue of approach towards the ideal state of maximum strength. Strangely enough, another line of argument suggests that the reverse of refinement might be effective in securing the desired result, in certain special cases. If a wire is required to withstand a simple tension, it seems that the best arrangement is that in which the strongest directions of all the molecules are parallel to the axis of the wire. This is equivalent to making the wire out of a single crystal. The theoretical tenacity would not be obtained, however, if the gliding planes made with the axis angles other than 0° C. and 90° C., as yield would occur.

If, in passing from the normal crystalline to the strong fine-grained state, the necessary orientation of the molecules were performed in accordance with some regular plan, the resulting configuration would possess some kind of symmetry, and the material might therefore exhibit crystalline properties. In cases where a substance exists in nature in several different crystalline forms, of which one is much stronger than the others, it may be that the strong modification is of the fine-grained type here considered. Thus, diamond may be a fine-grained modification of carbon. If this view is correct, it suggests that the transformation of carbon into diamond requires, firstly, the existence of conditions of temperature and pressure under which diamond has less potential energy than carbon ; and, secondly, the provision of means for causing relative rotation of the molecules. In the attempts which have so far been made in this direction, attention seems to have been concentrated on satisfying the former requirement, the possible existence of the latter one having been overlooked. The most obvious way of satisfying it, if the mechanical difficulties could be overcome, would appear to be the application of suitable shearing stresses in addition to the hydrostatic pressure.

## 11. *Application of the Theory to Liquids.*

A detailed discussion of the properties of liquids, in the light of the present theory, would scarcely fall within the scope of this paper. One prediction which has been made, however, and which has been verified experimentally, affords such a remarkable confirmation of the general theory that it is felt that no apology is necessary for introducing it here. Consider a solid composed of molecules whose attraction is a function of orientation, the molecules being arranged in groups, in accordance with the theory outlined in the preceding pages. If the temperature of this body be supposed to be increasing, it will be seen that at some temperature the kinetic reactions due to the thermal vibrations must overcome the minimum attractions of the molecules in each group. It is clear, therefore, that at this temperature the substance will be unable to withstand shearing stresses. At the same time it cannot vaporise, as the molecules must still be held together in chains by their maximum attractions. In other words, the transformation which has been discussed is simply the liquefaction of the solid.

This view of the phenomenon of melting indicates that the molecules of liquids are in general arranged in groups or chains, of a length comparable with that of the structure ascribed to solids in the preceding work, or, say, $10^4$ molecules.

If, therefore, a liquid be contained in a solid boundary which it wets, the ends of these chains may be expected to attach themselves to the solid; and if at any point the distance between the bounding walls is less than the length of the chains, some of the latter will attach themselves to both walls and hinder the free flow of the liquid and the relative movement, if any, of the boundaries. At such a point the liquid will act as a solid under any stress which is insufficient to break the chains.

This has been verified experimentally. The apparatus consisted of a polished steel ball 1 inch in diameter, and a block of hard tool steel containing a circular hole about 4 inches long. The hole was carefully ground, after hardening, to a diameter about 0·0001 inch greater, at its smallest part, than the diameter of the ball. When both were dry the ball passed freely through the hole. If, however, they were wetted with a liquid, considerable pressure was necessary to force the ball through. This resistance possessed the characteristic "stickiness" of solid friction, and was exactly the kind of resistance which would have been expected in forcing the dry ball through a hole which was too small for it.

To show that the resistance was a true "solid stress" and not due merely to viscosity, the apparatus was on one occasion left for a week, with the weight of the ball supported by the stress in the liquid (paraffin oil). The hole was vertical, so that there was no normal pressure between its surface and the surface of the ball. During this period no motion whatever could be detected.

It is essential to the success of these experiments that the ball and hole should be thoroughly wetted by the liquid. For this reason the liquids used have been chiefly paraffin oil and lubricating oils, but on one occasion the effect was obtained with water.

The present theory suggests a reason for the very low tensile strength of liquids. If a liquid is composed of a random aggregation of chains of molecules, it may reasonably be expected to contain regions of dimensions comparable with, but smaller than, the length of the chains, across which no chains run. Rupture of the liquid will evidently occur by the enlargement of these cavities. Now the tension, R, necessary to enlarge a spherical cavity of diameter, D, in a liquid of surface tension, T, is given by

$$R = 4T/D.$$

In the case of water, the tensile strength, R, is about 70 lbs. per sq. inch at ordinary temperatures, while T is about 0·00042 lbs. per inch. Hence the cavities, if spherical, must be at least 0·000024 inch in diameter. This is of the order indicated by the theory.

The foregoing conclusions are of especial interest in their relation to the theory of ROSENHAIN,* on which many of the properties of metals, and particularly "season cracking" under prolonged stress, are explained by supposing that the crystals are cemented together by very thin layers of amorphous material having the properties of an extremely viscous undercooled liquid. The experiments described above show that fluidity is not a property which can be ascribed *a priori* to such films. Hence if the view of ROSENHAIN and ARCHBUTT were to be definitely established, it would be necessary to regard it, not as a theory of season cracking in terms of the known properties of materials, but as a deduction of the properties of the intercrystalline layers from the phenomena of season cracking. Looked at in this way, it would be of extreme interest, for it would show that the molecular arrangement of the intercrystalline layers could not be of the coarse-grained type characteristic of the normal states of solids and liquids.

It is clear that the foregoing theory of liquids is not free from objection, and that in some respects it appears to be less satisfactory than existing theories. The most obvious objection is that it seems to be incompatible with accepted determination of the molecular weight of liquids. Since, however, these experiments are based ultimately on kinetic considerations, the author believes that this difficulty will not in fact arise unless the requisite bonds between the molecules of each group are found to be sufficiently strong to cause appreciable modification of the average molecular kinetic energy.

## 12. *Summary of Conclusions.*

(1) The ordinary hypothesis of rupture cannot be employed to predict the safe range of alternating stress which can be applied to metal having a scratched surface. The safe range of an unscratched test piece appears to be slightly less than the yield range, but if the surface is scratched the safe range may be several times the range which causes yield in the corners of the scratches.

(2) The "theorem of minimum potential energy" may be extended so as to be capable of predicting the breaking loads of elastic solids, if account is taken of the increase of surface energy which occurs during the formation of cracks.

(3) The breaking load of a thin plate of glass having in it a sufficiently long straight crack normal to the applied stress, is inversely proportional to the square root of the length of the crack. The maximum tensile stress in the corners of the crack is more than ten times as great as the tensile strength of the material, as measured in an ordinary test.

(4) The foregoing observation is in agreement with the known fact that the observed strength of materials is less than one-tenth of the strength deduced indirectly from physical data, on the assumption that the materials are isotropic. The observed

---

* W. ROSENHAIN and D. EWEN, "Intercrystalline Cohesion in Metals," 'J. Inst. Metals,' vol. 8 (1912); and W. ROSENHAIN and S. L. ARCHBUTT, "On the Intercrystalline Fracture of Metals under Prolonged Application of Stress (Preliminary Paper)," 'Roy. Soc. Proc.,' A, vol. 96 (1919).

strength is, in fact, no greater than it would be, according to the theory, if the test pieces contained cracks several thousand molecules long.

(5) It has been found possible to prepare rods and fibres of glass and fused quartz which have a tenacity of about one million pounds per square inch (approximately the theoretical strength) when tested in the ordinary way. The strength so observed diminishes spontaneously, however, to a lower steady value, which it reaches a few hours after the fibre has been prepared. This steady value depends on the diameter of the fibre. In the case of large rods it is the same as the ordinary tenacity, whereas in the finest fibres the strength diminishes but little from its initial high value. The relation between diameter and strength is of practically the same form for glass fibres as for metal wires.

(6) If it is assumed that intermolecular attraction is a function of the relative orientation of the attracting molecules, it is possible to construct a theory of all the phenomena mentioned in (3), (4) and (5) above: In the case of crystalline substances the theory also appears to explain yield and shearing fracture; elastic hysteresis; elastic afterworking; the fracture in tension of ductile materials and the flow of brittle materials under combined shearing stress and hydrostatic pressure; the drop in stress which occurs on the initiation of yield in ductile substances; fatigue failure under alternating stress; and the relatively slight effect of surface scratches on fatigue strength. In the case of non-crystalline materials the theory explains elastic afterworking and the great strength of short columns in compression.

(7) The theory shows that the application of the mathematical theory of homegeneous elastic solids to real substances may lead to error, unless the smallest material dimension involved, *e.g.*, the radius of curvature at the corner in the case of a scratch, is not many times the length of a crystal.

(8) It should be possible to raise the yield point of a crystalline substance by "refining" it, until at the ultimate limit of refinement the yield stress should be of the same order as the theoretical strength. It should also be possible similarly to increase the tenacity. Up to a certain stage the fatigue range should be unaffected by refining, but thereafter it should increase in the same degree as resistance to static stress.

(9) The theory requires that a thin film of liquid enclosed between solid boundaries which it wets should act as a solid. Experimental confirmation of this has been obtained.

In conclusion, the author desires to place on record his indebtedness to many past and present members of the staff of the Royal Aircraft Establishment for their valuable criticism and assistance, and also to Prof. C. F. JENKIN, at whose request the work on scratches was commenced.

[*Note.*—It has been found that the method of calculating the strain energy of a cracked plate, which is used in Section 3 of this paper, requires correction. The correction affects the numerical values of all quantities calculated from equations (6), (7), (8), (10), (11), (12) and (13), but not their order of magnitude. The main argument of the paper is therefore not impaired, since it deals only with the order of magnitude of the results involved, but some reconsideration of the experimental verification of the theory is necessary.]

# VII. *Reduction of Error by Linear Compounding.*

*By* W. F. SHEPPARD, *Sc.D.*, *LL.M.*

*Communicated by* E. T. WHITTAKER, *F.R.S.*

Received April 23,—Read June 24, 1920.

## CONTENTS.

Page
1. Introductory . . . . . . . . . . . . . . . . . . . . . . . . . . . . 200

CONJUGATE SETS.

2. Conjugate set . . . . . . . . . . . . . . . . . . . . . . . . . . . . 201
3. Relations between original set and conjugate set . . . . . . . . . . . 203
4. Two related pairs of conjugate sets . . . . . . . . . . . . . . . . 204
5. Sums as conjugates of differences . . . . . . . . . . . . . . . . . 205

REDUCTION OF ERROR (GENERAL).

6. General theorems . . . . . . . . . . . . . . . . . . . . . . . . . . 208
7. Notation: and particular values. . . . . . . . . . . . . . . . . . . 210
8. Improved values in terms of conjugates . . . . . . . . . . . . . . . 211
9. Mean products of error of improved values . . . . . . . . . . . . . 212
10. Expression in terms of a related set . . . . . . . . . . . . . . . 214
11. Special case of differences . . . . . . . . . . . . . . . . . . . . 215

RELATION OF "REDUCTION OF ERROR" TO "FITTING" (OF A POLYNOMIAL).

12. Standard system . . . . . . . . . . . . . . . . . . . . . . . . . . 215
13. Fitting by least squares . . . . . . . . . . . . . . . . . . . . . 216
14. Fitting by moments . . . . . . . . . . . . . . . . . . . . . . . . 219
15. Reduction of error . . . . . . . . . . . . . . . . . . . . . . . . 219
16. Difference of the two processes . . . . . . . . . . . . . . . . . . 219

SOME STEPS IN THE GENERAL SOLUTION.

17. Preliminary . . . . . . . . . . . . . . . . . . . . . . . . . . . . 220
18. Formula of progression . . . . . . . . . . . . . . . . . . . . . . 220
19. Mean products of error (alternative formula) . . . . . . . . . . . . 221

APPLICATION TO SELF-CONJUGATE SET.

20. Preliminary . . . . . . . . . . . . . . . . . . . . . . . . . . . . 222
21. Improved values . . . . . . . . . . . . . . . . . . . . . . . . . . 225
22. Extent of improvement (central differences) . . . . . . . . . . . . 230
23. Smoothing . . . . . . . . . . . . . . . . . . . . . . . . . . . . . 231

[Published October 22, 1920.

APPENDICES.

Page
APPENDIX I.—The correlation-determinant . . . . . . . . . . . . . . . . . . . 232
„ II.—Frequency of correlated errors . . . . . . . . . . . . . . . . . . . 233
„ III.—Improved advancing differences in terms of sums . . . . . . . . . . . 235
„ IV.—Formulæ in terms of $u$'s. . . . . . . . . . . . . . . . . . . . . 236

1. *Introductory.*—This paper is a development of two earlier papers,[*] which for brevity I call "Reduction" and "Fitting" respectively. The paper[†] immediately preceding "Fitting" is referred to as "Factorial Moments."

These earlier papers deal with two problems, which are closely connected and have the same solution. For both of them, the data are a set of quantities $u_0$, $u_1$, $u_2$, ... of the same kind, which we regard as representing certain true values $U_0$, $U_1$, $U_2$, ..., with errors $e_0$, $e_1$, $e_2$, ..., so that $u_r = U_r + e_r$. These errors may be independent or may be correlated in any way. The first problem is based on the assumption (which defines the class of cases we are dealing with) that the sequence of $U$'s is fairly regular, so that differences after those of a certain order, which we will call $j$, are negligible. This being so, we may alter any $u$, or any linear compound of the $u$'s, such as an interpolation-formula, by adding to it any linear compound of the negligible differences. (I use the term "linear compound" in preference to "linear function," since there is no consideration of functionality.) The problem is to find the value of the resulting sum when, by suitable choice of the coefficients in the added portion, the mean square of error of the sum is a minimum. This is the problem of "reduction of error." For the second problem it is assumed that $U_r$ is a polynomial in $r$ of degree $j$, and the problem is to find the coefficients in this polynomial by the method of least squares. This is the problem of "fitting."

The practical solution of these problems for the general case, in which the errors are correlated, is not easy. The particular case which is simple is that in which the errors all have the same mean square, which by a suitable choice of unit is taken to be 1, and the mean products of error are all 0. (In the previous papers I have called this system of errors the *standard* system; in the present paper the set of $u$'s which possesses this property is called a *self-conjugate* set.) In "Reduction" I gave the solution for this particular case in terms of central differences, and in "Fitting" I gave the solution in terms of advancing differences and of advancing and central sums, formed in a particular way. I also gave expressions in terms of the $u$'s, but these were rather complicated. It remained to obtain expressions for the mean squares of error of the new values, in order to compare them with those of the old values. In doing this I found that the whole of the work could be very much simplified by using certain general theorems, which applied not only to the special case of the standard system but also to the general case, and even to a still more general problem in which, in the one aspect, the reduction of error is effected by means of quantities which are not necessarily a set of differences, or in which, in the other aspect, $U_r$ is not necessarily a polynomial in $r$ of degree $j$ but is a linear compound, with coefficients to be determined, of any $j+1$ functions of $r$; and the present paper is mainly concerned with these general theorems, so that to a certain extent it supersedes the previous papers.

---

[*] "Reduction of Errors by means of Negligible Differences," 'Fifth International Congress of Mathematicians,' Cambridge, 1912, ii., 348–384; "Fitting of Polynomial by Method of Least Squares," 'Proceedings of the London Mathematical Society,' 2nd series, xiii., 97–108.

[†] "Factorial Moments in terms of Sums or Differences," 'Proceedings of the London Mathematical Society,' 2nd series, xiii., 81–96.

The abbreviations l.c., m.s.e., m.p.e., are used for linear compound, mean square of error, mean product of errors. The mean square of error of $A$ is denoted by $(A\ ;A)$, and the mean product of errors of $A$ and $B$ by $(A\ ;B)$ or $(B\ ;A)$. Other special notations used in the paper are the same as in the three papers mentioned at the beginning of this section, or are explained in §§ 3, 5 (iii.), 7, 17, and 20.

## CONJUGATE SETS.

2. *Conjugate Set.*—(i.) Let $A$, $B$, $C$, $D$, ... be a set of quantities, not necessarily all of the same kind, containing coexistent errors which are either independent or correlated in any way. For the purpose of the following investigations it is convenient to consider, in connexion with these quantities, another set of quantities, $G$, $H$, $J$, $K$, ... , equal to them in number and connected with them by the conditions that (1) each quantity of the second set is a l.c. of those of the first set, and (2) the m.p.e. of corresponding members of the two sets is 1 and that of members which do not correspond is 0. If we replace the quantities of the two sets by $A_0$, $A_1$, $A_2$, ... , and $G_0$, $G_1$, $G_2$, ... , we can express this latter condition by saying that m.p.e. of $G_r$ and $A_s = 0$ $(s \neq r)$ or $1$ $(s = r)$. The second set of quantities is said to be *conjugate* to the first.

(ii.) Let the member of the second set which corresponds to $C$ of the first set be $J$. To determine $J$, let us write

$$J = aA+bB+cC+dD+\ldots.$$

Then, denoting the m.p.e. of $A$ and $B$ by $(A\ ;B)$, condition (2) gives

$$\begin{aligned}
(A\ ;A)\,a+(A\ ;B)\,b+(A\ ;C)\,c+(A\ ;D)\,d+\ldots &= 0,\\
(B\ ;A)\,a+(B\ ;B)\,b+(B\ ;C)\,c+(B\ ;D)\,d+\ldots &= 0,\\
(C\ ;A)\,a+(C\ ;B)\,b+(C\ ;C)\,c+(C\ ;D)\,d+\ldots &= 1,\\
(D\ ;A)\,a+(D\ ;B)\,b+(D\ ;C)\,c+(D\ ;D)\,d+\ldots &= 0,\\
&\&c.
\end{aligned}$$

There are as many equations as there are coefficients $a$, $b$, $c$, $d$, ... ; and the values of these are thus uniquely determined.

(iii.) The values of $a$, $b$, $c$, ... as found from the above equations have as their denominator the determinant

$$\Theta \equiv \begin{vmatrix} (A\,;A) & (A\,;B) & (A\,;C) & (A\,;D) & \dots \\ (B\,;A) & (B\,;B) & (B\,;C) & (B\,;D) & \dots \\ (C\,;A) & (C\,;B) & (C\,;C) & (C\,;D) & \dots \\ (D\,;A) & (D\,;B) & (D\,;C) & (D\,;D) & \dots \\ \vdots & \vdots & \vdots & \vdots & \end{vmatrix}$$

There is therefore no conjugate set if this determinant is zero. The nature of the relations which in this case hold between the errors is considered in Appendix I., § 3.

(iv.) Since the members of the conjugate set are l.cc. of those of the original set, the converse also holds. Regrouping the equations which determine the coefficients, it will be seen that the original set is conjugate to the conjugate set; *i.e.*, that the two sets are conjugate to each other. The formulæ for the members of the original set in terms of those of the conjugate set are

$$\left.\begin{aligned} A &= (A\,;A)\,G+(A\,;B)\,H+(A\,;C)\,J+\dots, \\ B &= (B\,;A)\,G+(B\,;B)\,H+(B\,;C)\,J+\dots, \\ C &= (C\,;A)\,G+(C\,;B)\,H+(C\,;C)\,J+\dots, \\ &\quad \&c. \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (1)$$

These follow from the solution of the equations in (ii.), by the ordinary properties of determinants; or they may be obtained more simply by determining the coefficients of $G$, $H$, $J$, ... in each case from the second of the conditions stated in (i.).

(v). By means of these relations we can not only express any l.c. of the quantities of either set in terms of those of the other set, but we can also express any such l.c. in terms of particular quantities of one set and those of the conjugate set which correspond to the remaining quantities. We can, for instance, express any l.c. of $G$, $H$, $J$, $K$, ... in terms of $A$, $B$, $J$, $K$, ... by using the first two equations in (1) to determine $G$ and $H$ in terms of $A$, $B$, $J$, $K$, .... The results involve a certain determinant in the denominator; it is shown in Appendix I., § 4, that this is not zero if $\Theta$ is not zero.

(vi.) Two special cases may be mentioned :—

($a$) If the errors of $A$, $B$, $C$, $D$, ... form a standard system, *i.e.*, if the m.s.e. of each of the quantities is 1 and the m.p.e. of each pair of quantities is 0, the conjugate set is identical with the original set; and conversely. A set which is identical with the conjugate set will be called a *self-conjugate* set.

(*b*) If the m.p.e. of each pair of quantities of the original set is 0, but the m.ss.e. are not all 1, this is also the case for the conjugate set. The original set being $A$, $B$, $C$, ..., the quantities of the conjugate set are $A/(A\,;A)$, $B/(B\,;B)$, $C/(C\,;C)$, ...; and their m.ss.e. are $1/(A\,;A)$, $1/(B\,;B)$, $1/(C\,;C)$, ....

3. *Relations between Original Set and Conjugate Set.*—For expressing a member of either set in terms of the members of the other set, it is convenient to give them a linear order. We therefore denote the members of the original set by $\delta_0, \delta_1, \delta_2, \ldots \delta_l$ and those of the conjugate set by $\sigma_0, \sigma_1, \sigma_2, \ldots \sigma_l$. Also we write

$$\zeta_{r,t} \equiv \text{m.p.e. of } \delta_r \text{ and } \delta_t = \zeta_{t,r}, \quad \ldots \ldots \ldots \quad (2)$$

$$\eta_{r,t} \equiv \text{m.p.e. of } \sigma_r \text{ and } \sigma_t = \eta_{t,r}. \quad \ldots \ldots \ldots \quad (3)$$

(i.) The condition of conjugacy is that ($r = 0, 1, 2, \ldots l$; $t = 0, 1, 2, \ldots l$)

$$\text{m.p.e. of } \delta_r \text{ and } \sigma_t = 0\,(r \neq t) \quad \text{or} \quad 1\,(r = t). \quad \ldots \ldots \quad (4)$$

(ii.) The expression for $\delta_r$ in terms of the $\sigma$'s is (*cf.* § 2 (iv.))

$$\delta_r = \zeta_{r,0}\sigma_0 + \zeta_{r,1}\sigma_1 + \zeta_{r,2}\sigma_2 + \ldots + \zeta_{r,l}\sigma_l. \quad \ldots \ldots \ldots \quad (5)$$

[For, if we write

$$\delta_r = a_0\sigma_0 + a_1\sigma_1 + a_2\sigma_2 + \ldots + a_l\sigma_l,$$

then (2) and (4) give

$$\begin{aligned} \zeta_{r,t} &= \text{m.p.e. of } \delta_t \text{ and } a_0\sigma_0 + a_1\sigma_1 + \ldots + a_l\sigma_l \\ &= a_t.] \end{aligned}$$

(iii.) Similarly the expression for $\sigma_t$ in terms of the $\delta$'s is

$$\sigma_t = \eta_{0,t}\delta_0 + \eta_{1,t}\delta_1 + \eta_{2,t}\delta_2 + \ldots + \eta_{l,t}\delta_l. \quad \ldots \ldots \ldots \quad (6)$$

(iv.) The relations between the $\zeta$'s and the $\eta$'s are easily deduced from (5) and (6). If we write

$$Z \equiv \begin{vmatrix} \zeta_{0,0} & \zeta_{0,1} & \zeta_{0,2} & \cdots & \zeta_{0,l} \\ \zeta_{1,0} & \zeta_{1,1} & \zeta_{1,2} & \cdots & \zeta_{1,l} \\ \zeta_{2,0} & \zeta_{2,1} & \zeta_{2,2} & \cdots & \zeta_{2,l} \\ \vdots & \vdots & \vdots & & \vdots \\ \zeta_{l,0} & \zeta_{l,1} & \zeta_{l,2} & \cdots & \zeta_{l,l} \end{vmatrix}, \quad \ldots \ldots \ldots \quad (7)$$

$$Z_{p,q} \equiv \text{cofactor of } \zeta_{p,q} \text{ in } Z = Z_{q,p}, \quad \ldots \ldots \ldots \quad (8)$$

$$H \equiv \begin{vmatrix} \eta_{0,0} & \eta_{0,1} & \eta_{0,2} & \cdots & \eta_{0,l} \\ \eta_{1,0} & \eta_{1,1} & \eta_{1,2} & \cdots & \eta_{1,l} \\ \eta_{2,0} & \eta_{2,1} & \eta_{2,2} & \cdots & \eta_{2,l} \\ \vdots & \vdots & \vdots & & \vdots \\ \eta_{l,0} & \eta_{l,1} & \eta_{l,2} & \cdots & \eta_{l,l} \end{vmatrix}, \quad \dots \dots \quad (9)$$

$$H_{p,q} \equiv \text{cofactor of } \eta_{p,q} \text{ in } H = H_{q,p}, \quad \dots \dots \quad (10)$$

then

$$\eta_{p,q} = Z_{p,q}/Z, \quad \dots \dots \quad (11)$$

$$\zeta_{p,q} = H_{p,q}/H, \quad \dots \dots \quad (12)$$

$$HZ = 1. \quad \dots \dots \quad (13)$$

(v.) The assumption that there is a conjugate set implies (*cf.* § 2 (iii.)) that $Z$ is not $= 0$. It follows from (13) that $H$ is not $= 0$. It also follows (see Appendix I., § 4 ($b$)) that none of the principal minors of $Z$ or of $H$ are $= 0$.

4. *Two Related Pairs of Conjugate Sets.*—(i.) Suppose that there is another set of $l+1$ quantities $u_0, u_1, u_2, \dots u_l$, connected with the $\delta$'s by the linear relations ($r = 0, 1, 2, \dots l$)

$$u_r = (r_0)\,\delta_0 + (r_1)\,\delta_1 + (r_2)\,\delta_2 + \dots + (r_l)\,\delta_l. \quad \dots \dots \quad (14)$$

Then, by the condition of conjugacy of the $\delta$'s and the $\sigma$'s,

$$(r_t) = \text{m.p.e. of } u_r \text{ and } \sigma_t. \quad \dots \dots \quad (15)$$

Let the set conjugate to $u_0, u_1, u_2, \dots u_l$ be $y_0, y_1, y_2, \dots y_l$. Then there are linear relations between the $y$'s and the $u$'s and between the $\sigma$'s and the $\delta$'s, and therefore also, by (14), between the $y$'s and the $\sigma$'s. To find the $\sigma$'s in terms of the $y$'s, we write (15) in the form

$$(r_t) = \text{m.p.e. of } \sigma_t \text{ and } u_r\,;$$

and we see that ($t = 0, 1, 2, \dots l$)

$$\sigma_t = (0_t)\,y_0 + (1_t)\,y_1 + (2_t)\,y_2 + \dots + (l_t)\,y_l. \quad \dots \dots \quad (16)$$

(ii.) Similarly, if the expression for the $\delta$'s in terms of the $u$'s is ($s = 0, 1, 2, \dots l$)

$$\delta_s = \{s_0\}\,u_0 + \{s_1\}\,u_1 + \{s_2\}\,u_2 + \dots + \{s_l\}\,u_l, \quad \dots \dots \quad (17)$$

where, by (14),

$$(r_0)\,\{0_t\} + (r_1)\,\{1_t\} + \dots + (r_l)\,\{l_t\} = 0\,(r \neq t) \quad \text{or} \quad 1\,(r = t), \quad \dots \quad (18)$$

$$\{s_0\}\,(0_t) + \{s_1\}\,(1_t) + \dots + \{s_l\}\,(l_t) = 0\,(r \neq t) \quad \text{or} \quad 1\,(r = t), \quad \dots \quad (19)$$

then

$$\{s_t\} = \text{m.p.e. of } \delta_s \text{ and } y_t, \quad \dots \dots \quad (20)$$

$$y_t = \{0_t\}\,\sigma_0 + \{1_t\}\,\sigma_1 + \{2_t\}\,\sigma_2 + \dots + \{l_t\}\,\sigma_l. \quad \dots \dots \quad (21)$$

(iii.) The above relations can be expressed diagrammatically, thus :—

$$\begin{array}{c|ccccc|c}
\vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\
\delta_3 & \times & \times & \times & \times & \ldots & \sigma_3 \\
\delta_2 & \times & \times & \times & \times & \ldots & \sigma_2 \\
\delta_1 & \times & \times & \times & \times & \ldots & \sigma_1 \\
\delta_0 & \times & \times & \times & \times & \ldots & \sigma_0 \\
\hline
 & u_0 & u_1 & u_2 & u_3 & \ldots & \\
\hline
 & y_0 & y_1 & y_2 & y_3 & \ldots &
\end{array}$$

The crosses represent the ( ) coefficients if they are the coefficients of the $\delta$'s in the $u$'s and of the $y$'s in the $\sigma$'s, and the { } coefficients in the converse case.

(iv.) Similarly, if we write ($r = 0, 1, 2, \ldots l$ ; $t = 0, 1, 2, \ldots l$)

$$[r_t] \equiv \text{m.p.e. of } y_r \text{ and } \sigma_t, \qquad (22)$$

then

$$y_r = [r_0]\,\delta_0 + [r_1]\,\delta_1 + [r_2]\,\delta_2 + \ldots + [r_l]\,\delta_l. \qquad (23)$$

$$\sigma_t = [0_t]\,u_0 + [1_t]\,u_1 + [2_t]\,u_2 + \ldots + [l_t]\,u_l. \qquad (24)$$

5. *Sums as Conjugates of Differences.*—The cases of importance are those in which the $\delta$'s are successive differences of the $u$'s. It will be found that in these cases the $\sigma$'s are l.cc. of successive sums of the $y$'s.

(i.) Let the $\delta$'s be the advancing differences of the $u$'s, *i.e.*

$$\delta_0 = u_0,\ \delta_1 = \Delta u_0,\ \ldots\ \delta_r = \Delta^r u_0,\ \ldots.$$

Then the diagram for the ( ) coefficients is

$$\begin{array}{c|ccccc|c}
\vdots & \vdots & \vdots & \vdots & \vdots & & \vdots \\
\delta_3 & 0 & 0 & 0 & 1 & \ldots & \sigma_3 \\
\delta_2 & 0 & 0 & 1 & 3 & \ldots & \sigma_2 \\
\delta_1 & 0 & 1 & 2 & 3 & \ldots & \sigma_1 \\
\delta_0 & 1 & 1 & 1 & 1 & \ldots & \sigma_0 \\
\hline
 & u_0 & u_1 & u_2 & u_3 & \ldots & \\
\hline
 & y_0 & y_1 & y_2 & y_3 & \ldots &
\end{array}$$

so that

$$\sigma_0 = y_0+y_1+y_2+y_3+\ldots+y_l,$$

$$\sigma_1 = y_1+2y_2+3y_3+4y_4+\ldots+ly_l,$$

$$\sigma_2 = y_2+3y_3+6y_4+10y_5+\ldots+\tfrac{1}{2}l\,(l-1)\,y_l,$$

and, generally,

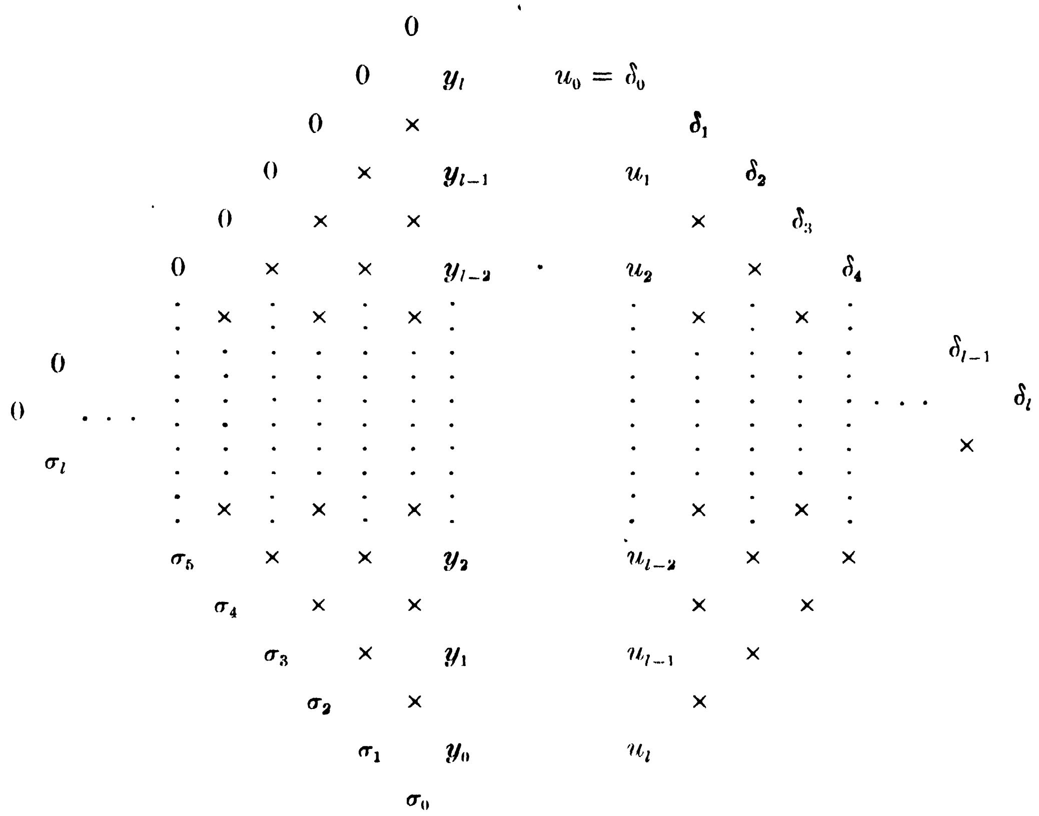
$$\sigma_f = y_f+(f+1,\,1)\,y_{f+1}+(f+2,\,2)\,y_{f+2}+\ldots+(l,\,l-f)\,y_l$$

$$= \sum_{q=f}^{q=l} (q,f)\,y_q\,; \qquad (25)$$

or, in the notation of "Fitting," § 4, and "Factorial Moments,"

$$\sigma_f = \Sigma''^{f+1}y_f. \qquad (26)$$

The $\sigma$'s can be obtained by successive summations of the $y$'s in reverse order, *i.e.* from $y_l$ to $y_0$, as shown in the following diagram, in which the crosses represent entries in a sum- or difference-table :—



(ii.) Let the $\delta$'s be the central differences of the $u$'s. Then it will be found in the same way that

(*a*) If the $u$'s are $u_0$, $u_1$, $u_2$, ... $u_{2n}$, so that

$$\delta_0 = u_n,\ \delta_1 = \mu\delta u_n,\ \delta_2 = \delta^2 u_n,\ \delta_3 = \mu\delta^3 u_n,\ \ldots,$$

then

$$\sigma_{2h} = \sum_{r=-n}^{r=n} [r, 2h)\, y_{n+r}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (27)$$

$$\sigma_{2h-1} = \sum_{r=-n}^{r=n} (r, 2h-1]\, y_{n+r}; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (28)$$

(*b*) If the $u$'s are $u_0, u_1, u_2, \ldots u_{2n-1}$, so that

$$\delta_0 = \mu u_{n-\frac{1}{2}},\ \delta_1 = \delta u_{n-\frac{1}{2}},\ \delta_2 = \mu\delta^2 u_{n-\frac{1}{2}},\ \delta_3 = \delta^3 u_{n-\frac{1}{2}}, \ldots,$$

then

$$\sigma_{2h-1} = \sum_{r=-n+1}^{r=n} [r-\tfrac{1}{2}, 2h-1)\, y_{n+r-1}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (29)$$

$$\sigma_{2h-2} = \sum_{r=-n+1}^{r=n} (r-\tfrac{1}{2}, 2h-2]\, y_{n+r-1}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (30)$$

(iii.) The values given by (25)–(30) may be expressed in terms of successive sums by the formulæ given in "Factorial Moments." The notation, however, can be simplified. Suppose that we have a set of quantities $\ldots F_0, F_1, F_2, \ldots$ corresponding to values $\ldots 0, 1, 2, \ldots$ of some variable, and that we form the table of successive differences (and also, if we like, of successive sums) of the $F$'s. Then the Lagrangian formula for $F_\theta$ in terms of $F_p, F_{p+1}, \ldots F_{p+t}$, which can be expressed in a good many different ways, may be regarded as the formula for it in terms of the whole (unlimited) set of differences (and sums) which form a triangle with its apex at $\Delta^t F_p$; and we can denote it by $L\{F_\theta; \Delta^t F_p\}$. With this notation, the above results may be written

(25)
$$\sigma_f = \left[L\{(-)^f \Sigma^{f+1} y_f; \Sigma y_t\}\right]_{t=0}^{t=l+1}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (31)$$

(27)
$$\sigma_{2h} = \left[L\{\mu\sigma^{2h+1} y_n; \sigma y_{n+t}\}\right]_{t=\ \ n-\frac{1}{2}}^{t=n+\frac{1}{2}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (32)$$

(28)
$$\sigma_{2h-1} = \left[L\{-\sigma^{2h} y_n; \sigma y_{n+t}\}\right]_{t=-n-\frac{1}{2}}^{t=n+\frac{1}{2}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (33)$$

(29)
$$\sigma_{2h-1} = \left[L\{-\mu\sigma^{2h} y_{n-\frac{1}{2}}; \sigma y_{n+t-1}\}\right]_{t=-n+\frac{1}{2}}^{t=n+\frac{1}{2}}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (34)$$

(30)
$$\sigma_{2h-2} = \left[L\{\sigma^{2h-1} y_{n-\frac{1}{2}}; \sigma y_{n+t-1}\}\right]_{t=-n+\frac{1}{2}}^{t=n+\frac{1}{2}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (35)$$

The $L$ is distributive as regards the first member inside the $\{\ \}$; *e.g.*, in the case of (31),

$$A\sigma_2 + B\sigma_3 = \left[L\{A\Sigma^3 y_2 - B\Sigma^4 y_3; \Sigma y_t\}\right]_{t=0}^{t=l+1}.$$

(iv.) More generally, suppose that the $\delta$'s are the successive differences of the $u$'s according to any system of differences; by which we mean that $\delta_s$ is either a definite difference of the $u$'s of order $s$ (the $u$'s themselves being differences of order 0) or a l.c. of such differences. Then $(r_t)$ of (14) is a polynomial in $r$ of degree $t$, and $\sigma_t$ is of the form $\sum_{q=0}^{q=t} \phi_t(q)\,y_q$, where $\phi_t(q)$ is some polynomial in $q$ of degree $t$. It follows that any l.c. of $\sigma_0, \sigma_1, \sigma_2, \dots \sigma_t$ is also of this form.

(v.) If we denote the $f^{\text{th}}$ moment of the $y$'s by $M_f$, then $M_f$ is of the form $\sum_{q=0}^{q=m} \phi_f(q)\,y_q$. Hence $\sigma_t$ is a l.c. of $M_0, M_1, M_2, \dots M_t$; and $M_t$ is a l.c. of $\sigma_0, \sigma_1, \sigma_2, \dots \sigma_t$. More generally, any l.c. of $\sigma_0, \sigma_1, \sigma_2, \dots \sigma_t$ is a l.c. of $M_0, M_1, M_2, \dots M_t$; and conversely.

## REDUCTION OF ERROR (GENERAL).

6. *General Theorems.*—Let $A$, $B$, $C$, ... $P$, $Q$, $R$, ... be a set of quantities as in § 2, but all of the same kind. If

$$w \equiv aA+bB+cC+\dots \text{ (with or without terms in } P,\ Q,\ R, \dots),$$

$$x = w+pP+qQ+rR+\dots,$$

where $a$, $b$, $c$, ... are fixed and $p$, $q$, $r$, ... are arbitrary, and if we choose $p$, $q$, $r$, ... so as to make the m.s.e. of $x$ a minimum, the resulting value of $x$ is called the *improved value of* w, *using* P, Q, R, ... *as auxiliaries.* The following are general theorems; some are quite elementary, but it is convenient to state them here. The specially important theorems are (III.) and (XIII.). The assumption mentioned under (VI.) should be noted. If strict proofs of (I.) and (II.) are required, the method should be that of Appendix I., § 2.

(I.) *The m.p.e. of* A *and any l.c. of* A, B, C, ... *is the same l.c. of the m.pp.e. of* A *and* A, B, C, ... [*i.e.*, m.p.e. of $A$ and $aA+bB+cC+\dots$ is $a(A\ ;\ A)+b(A\ ;\ B)+c(A\ ;\ C)+\dots$].

(II.) *The m.s.e. of any l.c. of* A, B, C, ..., *or the m.p.e. of any two such l.cc., is found by squaring the former or multiplying the latter and replacing squares and products by the corresponding m.ss.e. and m.pp.e.* [*i.e.*, m.s.e. of $aA+bB+cC+\dots$ $= a^2(A\ ;\ A)+2ab(A\ ;\ B)+b^2(B\ ;\ B)+2ac(A\ ;\ C)+2bc(B\ ;\ C)+c^2(C\ ;\ C)+\dots$, and similarly for m.p.e. of $aA+bB+cC+\dots$ and $a'A+b'B+c'C+\dots$].

(III.) *If the improved value of* A, *using certain auxiliaries, is* $\mathrm{A}+\alpha$, *then the m.p.e. of* $\mathrm{A}+\alpha$ *and each of the auxiliaries or* $\alpha$ *or any other l.c. of the auxiliaries is zero.* [Let the auxiliaries be $P$, $Q$, $R$, ..., and let $A+\alpha = A+pP+qQ+rR+\dots$. Then the m.s.e. of $A+(p+\theta)P+qQ+rR+\dots (= A+\alpha+\theta P)$ is $(A+\alpha\ ;\ A+\alpha)+2\theta(A+\alpha\ ;\ P)+\theta^2(P\ ;\ P)$. In order that this may be a minimum for $\theta = 0$, $(A+\alpha\ ;\ P)$ must be zero. Similarly for $(A+\alpha\ ;\ Q)$, $(A+\alpha\ ;\ R)$, .... This proves the first part of the theorem; the second then follows from (I.).] Hence

(IV.) *If the improved values of* A *and of* B, *using in each case the same set of auxiliaries, are* $\mathrm{A}+\alpha$ *and* $\mathrm{B}+\beta$, *the m.pp.e. of* $\mathrm{A}+\alpha$ *and* $\mathrm{B}+\beta$, *of* $\mathrm{A}+\alpha$ *and* B, *and of* A *and* $\mathrm{B}+\beta$, *are all equal;* and

(V.) *If the improved value of* A, *using certain auxiliaries, is* $\mathrm{A}+\alpha$, *and that of* B, *using some only of these, is* $\mathrm{B}+\beta$, *the m.p.e. of* $\mathrm{A}+\alpha$ *and* $\mathrm{B}+\beta$ *is equal to that of* $\mathrm{A}+\alpha$ *and* B.

(VI.) *If the improved value of* A, *using* P, Q, R, ... *as auxiliaries, is* $\mathrm{A}+\mathrm{pP}+\mathrm{qQ}+\mathrm{rR}+\ldots$, *the values of* p, q, r, ... *are given by a set of linear relations between* p, q, r, ... *and the m.pp.e. of* A *and* P, Q, R, .... [The equations are given by (III.), viz., since $(A+pP+qQ+rR+\ldots\,;\,P)=0$, &c.,

$$
\begin{aligned}
(A\,;\,P)+p\,(P\,;\,P)+q\,(Q\,;\,P)+r\,(R\,;\,P)+\ldots &= 0,\\
(A\,;\,Q)+p\,(P\,;\,Q)+q\,(Q\,;\,Q)+r\,(R\,;\,Q)+\ldots &= 0,\\
(A\,;\,R)+p\,(P\,;\,R)+q\,(Q\,;\,R)+r\,(R\,;\,R)+\ldots &= 0,\\
&\text{\&c.}
\end{aligned}
$$

It is assumed that the determinant

$$
\begin{vmatrix}
(P\,;\,P) & (Q\,;\,P) & (R\,;\,P) & \ldots\\
(P\,;\,Q) & (Q\,;\,Q) & (R\,;\,Q) & \ldots\\
(P\,;\,R) & (Q\,;\,R) & (R\,;\,R) & \ldots\\
\vdots & \vdots & \vdots &
\end{vmatrix}
$$

is not zero ; *i.e.* (see § 2 (iii.)) that there is a set conjugate to $P$, $Q$, $R$, ....]

(VII.) *For any particular set of auxiliaries there is one and only one improved value of* A. [This follows from the fact that the equations in (VI.), on the assumption there stated, have one and only one solution.] Hence we get the converse of (III.) :—

(VIII.) *If* x *is the sum of* w *and a l.c. of* P, Q, R, ..., *and if the m.p.e. of* x *and each of* P, Q, R, ... *is zero, then* x *is the improved value of* w, *using* P, Q, R, ... *as auxiliaries.* As a particular case :—

(IX.) *If the m.p.e. of* w *and each of* P, Q, R, ... *is zero, the improved value of* w, *using* P, Q, R, ... *as auxiliaries, is the same as the original value.*

(X.) *The improved value of* P, *using* P, Q, R, ... *as auxiliaries, is* $\mathrm{P}-\mathrm{P}=0$. [This follows either from (VIII.) or from taking $A=P$ in the equations in (VI.).] Hence

(XI.) *If* w *is a l.c. of* A, B, C, ... P, Q, R, ..., *the improved value of* w, *using* P, Q, R, ... *as auxiliaries, is the same as that of the quantity obtained by adding to* w *any l.c. of* P, Q, R, ... ; and, conversely :—

(XII.) *If the improved values of* w *and of* w′, *using in each case the same set of auxiliaries, are identical, then* w *and* w′ *either are identical or differ by a l.c. of the auxiliaries.*

(XIII.) *If the improved values of* A, B, C, ..., *using in each case the same set of* *auxiliaries, are* $\mathrm{A}+\alpha$, $\mathrm{B}+\beta$, $\mathrm{C}+\gamma, \ldots$, *the improved value of any l.c. of* A, B, C, ..., *using these auxiliaries, is the same l.c. of* $\mathrm{A}+\alpha$, $\mathrm{B}+\beta$, $\mathrm{C}+\gamma, \ldots$. [Let the l.c. of $A, B, C, \ldots$ be $w \equiv aA+bB+cC+\ldots$. We want to prove that $x \equiv a(A+\alpha)+b(B+\beta)+c(C+\gamma)+\ldots$ is its improved value. We can do this in either of two ways:

(i.) By (III.), the m.p.e. of $x$ and each of the auxiliaries $P, Q, R, \ldots$ is zero; and $x$ differs from $w$ by a l.c. of $P, Q, R, \ldots$. Hence, by (VIII.), $x$ is the improved value of $w$, using $P, Q, R, \ldots$ as auxiliaries.

(ii.) A more direct proof follows from the linearity of the equations mentioned in (VI.). It is not necessary to set out the proof here.]

(XIV.) *If* A, B, ... C, D, ... P, Q, ... R, S, ... *fall into two classes* A, B, ... P, Q, ... *and* C, D, ..., R, S, ..., *such that the m.p.e. of each member of the one class and each member of the other class is zero, then the improved value of a l.c. of any of the members, using* P, Q, ... R, S, ... *as auxiliaries, is to be found by taking the two classes separately, i.e.*, by using $P, Q, \ldots$ as auxiliaries for the terms in $A, B, \ldots P, Q, \ldots$, and $R, S, \ldots$ as auxiliaries for the terms in $C, D, \ldots R, S, \ldots$. [For the m.s.e. of $aA+bB+\ldots+cC+dD+\ldots+pP+qQ+\ldots+rR+sS+\ldots$ is the sum of those of $aA+bB+\ldots+pP+qQ+\ldots$ and $cC+dD+\ldots+rR+sS+\ldots$, since the m.p.e. of these latter is zero; we cannot reduce the m.s.e. of the first of them by adding terms in $R, S, \ldots$, or that of the second of them by adding terms in $P, Q, \ldots$; and the result is therefore the same as if we considered them separately.]

(XV.) *If the improved value of* w, *using* P, Q, R, ... *as auxiliaries, is* $\mathrm{x}=\mathrm{w}+\mathrm{pP}+$*terms in* Q, R, ..., *this is also the improved value of* $\mathrm{w}+\mathrm{pP}$, *using* Q, R, ... *as auxiliaries.* [For $x$ differs from $w+pP$ by terms in $Q, R, \ldots$, and the m.p.e. of $x$ and each of $Q, R, \ldots$ is zero.] This can be stated more generally as follows:—

(XVI.) *If the improved value of* A, *using a set of auxiliaries* S, *is* $\mathrm{A}+\alpha$, *and if we divide* S *into two sets,* $\mathrm{S}_1$ *and* $\mathrm{S}_2$, *and the corresponding parts of* $\alpha$ *are* $\alpha_1$ *and* $\alpha_2$, *then* $\mathrm{A}+\alpha$ *is the improved value of* $\mathrm{A}+\alpha_1$, *using* $\mathrm{S}_2$ *as auxiliaries.* [We may take $\ldots P, Q$ to be $S_1$, and $R, \ldots$ to be $S_2$. The theorem states that, if the improved value of $A$, using $\ldots P, Q, R, \ldots$, is $A+\ldots+pP+qQ+rR+\ldots$, this is also the improved value of $A+\ldots+pP+qQ$, using $R, \ldots$.]

(XVII.) The following corollaries of (III.) may be noticed, though we shall not require them. If the improved values of $A$ and of $B$, using in each case the same set of auxiliaries, are $A+\alpha$ and $B+\beta$, then

$$(1)\ (A+\alpha\,;\ A+\alpha)=(A\,;\ A)-(\alpha\,;\ \alpha)$$

and

$$(2)\ (A+\alpha\,;\ B+\beta)=(A\,;\ B)-(\alpha\,;\ \beta).$$

7. *Notation: and Particular Values.*—(i.) It will now be convenient to adopt a linear arrangement of the quantities we are dealing with, and we therefore replace

$A$, $B$, $C$, ... $P$, $Q$, $R$, ... by $\delta_0$, $\delta_1$, $\delta_2$, ... $\delta_{j+1}$, $\delta_{j+2}$, ... $\delta_l$. The order is quite arbitrary, so far as any general theorems are concerned; but it will usually be convenient to place the auxiliaries last. If, for instance, we are using all but $j+1$ as auxiliaries, we denote those not so used by $\delta_0$, $\delta_1$, $\delta_2$, ... $\delta_j$, and the auxiliaries by $\delta_{j+1}$, $\delta_{j+2}$, ... $\delta_l$; the improved values are then denoted by $(\ )_j$.

(ii.) We use the following notation:—

$$(\epsilon_f)_j \equiv \text{improved value of } \delta_f, \text{ using } \delta\text{'s after } \delta_j;$$

$$\mathrm{E}_j \equiv (\epsilon_j)_j = \text{improved value of } \delta_j, \text{ using all subsequent } \delta\text{'s};$$

$$(\lambda_{f,g})_j \equiv \text{m.p.e. of } (\epsilon_f)_j \text{ and } (\epsilon_g)_j;$$

$$\Lambda_j \equiv (\lambda_{j,j})_j = \text{m.s.e. of } \mathrm{E}_j.$$

Where there is no doubt as to the $\delta$'s used as auxiliaries, $(\epsilon_f)_j$ and $(\lambda_{f,g})_j$ can be replaced by $\epsilon_f$ and $\lambda_{f,g}$.

(iii.) By (X.) of §6—

$$(\epsilon_f)_j = 0 \text{ if } f > j; \quad \ldots \quad (36)$$

$$(\lambda_{f,g})_j = 0 \text{ if } f > j \text{ or } g > j. \quad \ldots \quad (37)$$

(iv.) By (IV.)—

$$\left.\begin{aligned}(\lambda_{f,g})_j &= \text{m.p.e. of } (\epsilon_f)_j \text{ and } \delta_g \\ &= \text{m.p.e. of } \delta_f \text{ and } (\epsilon_g)_j\end{aligned}\right\}; \quad \ldots \quad (38)$$

$$\Lambda_j = \text{m.p.e. of } \mathrm{E}_j \text{ and } \delta_j. \quad \ldots \quad (39)$$

8. *Improved Values in terms of Conjugates.*—In "Fitting" I have given some formulæ for improved values in terms of sums. These may be regarded as derivable from a general theorem relating to the expression of improved values in terms of members of the conjugate set. The theorem is given by (XIX.) and (XX.) below; (XVIII.) is a particular case.

(i.) Take any one of the $\delta$'s as $\delta_0$. By (6)—

$$\sigma_0 = \eta_{0,0}\delta_0 + \eta_{1,0}\delta_1 + \ldots + \eta_{l,0}\delta_l.$$

Hence $\sigma_0/\eta_{0,0}$ differs from $\delta_0$ by a l.c. of the other $\delta$'s. Also the m.p.e. of $\sigma_0/\eta_{0,0}$ and each of these other $\delta$'s is zero. It follows from (VIII.) of §6 that $\sigma_0/\eta_{0,0}$ is the improved value of $\delta_0$, using the other $\delta$'s as auxiliaries. The m.s.e. of this improved value is $\eta_{0,0}/(\eta_{0,0})^2 = 1/\eta_{0,0}$. Hence—

(XVIII.) *The improved value of any member of the original set, taking all the other members as auxiliaries, is the product of the corresponding member of the conjugate set by a constant; this constant being = the m.s.e. of the improved value.*

(ii.) The first part of the more general theorem is:—

(XIX.) *The improved value of any l.c. of a set of quantities, using those after the first* j+1 *as auxiliaries, is a l.c. of the first* j+1 *of the conjugate set.*

For, if $w$ is a l.c. of $\delta_0, \delta_1, \delta_2, \ldots \delta_l$, we can (see §2 (v.)) express it as a l.c. of $\sigma_0, \sigma_1, \sigma_2, \ldots \sigma_j, \delta_{j+1}, \delta_{j+2}, \ldots \delta_l$. Let the result be $(\Sigma)+(\Delta)$, where $(\Sigma)$ is a l.c. of $\sigma_0, \sigma_1, \sigma_2, \ldots \sigma_j$, and $(\Delta)$ is a l.c. of $\delta_{j+1}, \delta_{j+2}, \ldots \delta_l$. Then $(\Sigma)$ differs from $w$ by a l.c. of these latter $\delta$'s, and the m.p.e. of $(\Sigma)$ and each of these $\delta$'s is 0; hence, by (VIII.), $(\Sigma)$ is the improved value of $w$, using these $\delta$'s as auxiliaries.

(iii.) Further—

(XX.) *The coefficients of the $\sigma$'s in the improved value of the l.c. are the m.pp.e. of this improved value and the corresponding $\delta$'s.*

For, if the improved value of $w$ is $x$, and we write

$$x = b_0\sigma_0+b_1\sigma_1+b_2\sigma_2+\ldots+b_j\sigma_j,$$

then, by the condition of conjugacy of the $\sigma$'s and the $\delta$'s,

$$\text{m.p.e. of } x \text{ and } \delta_f = b_f.$$

(iv.) This would give us a solution of the problem of finding the improved value, if we could find the m.pp.e. Ordinarily, $w$ is or can be expressed in terms of the $\delta$'s, and we do not find its improved value independently, but deduce it from those of the $\delta$'s up to $\delta_j$. The improved value of $\delta_h$ is, by (iii),

$$(\epsilon_h)_j = (\lambda_{0,h})_j\sigma_0+(\lambda_{1,h})_j\sigma_1+\ldots+(\lambda_{j,h})_j\sigma_j; \quad . \quad . \quad . \quad . \quad . \quad . \quad (40)$$

and the m.pp.e. that we really require are therefore the values of $(\lambda_{f,h})_j$. With regard to this, see §9.

(v.) As the converse of (XIX.) it may be noted that—

(XXI.) *A quantity of the conjugate set, or a l.c. of such quantities, cannot be improved by means of the non-corresponding quantities of the original set; e.g.*, a l.c. of $\sigma_3$ and $\sigma_4$ cannot be improved by using the $\delta$'s, other than $\delta_3$ and $\delta_4$, as auxiliaries. This follows from (IX.) of §6, since the m.p.e. of $\delta_r$ and $\sigma_s$ is 0 unless $r = s$.

(vi.) If, as in § 5, there are related conjugate sets of $u$'s and $y$'s, and the $\delta$'s are the differences of the $u$'s, it follows from § 5 (v.) that $(\Sigma)$ in (ii.) above is a l.c. of the moments of the $y$'s up to the $j^{\text{th}}$. (XIX.) is therefore a generalisation of the theorem, for a self-conjugate set, that the improved values are l.cc. of the moments; and, in fact, it explains the appearance of the moments in this connexion.

9. *Mean Products of Error of Improved Values.*—(i.) We have found, in § 8 (iv.), that

$$(\epsilon_h)_j = (\lambda_{0,h})_j\sigma_0 + (\lambda_{1,h})_j\sigma_1+\ldots+(\lambda_{j,h})_j\sigma_j.$$

To obtain the $\lambda$'s, we introduce the condition that this shall differ from $\delta_h$ by a l.c. of $\delta$'s after $\delta_j$.

(ii.) Substituting for $\sigma_0$, $\sigma_1$, $\sigma_2$, ... from (6), this condition gives

$$\begin{aligned}&(\lambda_{0,h})_j\,(\eta_{0,0}\delta_0+\eta_{1,0}\delta_1+\ldots+\eta_{l,0}\delta_l)\\ &+(\lambda_{1,h})_j\,(\eta_{0,1}\delta_0+\eta_{1,1}\delta_1+\ldots+\eta_{l,1}\delta_l)\\ &+(\lambda_{2,h})_j\,(\eta_{0,2}\delta_0+\eta_{1,2}\delta_1+\ldots+\eta_{l,2}\delta_l)\\ &+\ldots\\ &+(\lambda_{j,h})_j\,(\eta_{0,j}\delta_0+\eta_{1,j}\delta_1+\ldots+\eta_{l,j}\delta_l)\\ &=\delta_h+\text{ terms in }\delta_{j+1},\ \delta_{j+2},\ \ldots\ \delta_l.\end{aligned}$$

Equating the coefficients of $\delta_0$, $\delta_1$, $\delta_2$, ... $\delta_j$, we find that $(f=0,\ 1,\ 2,\ \ldots,j)$

$$\eta_{f,0}(\lambda_{0,h})_j+\eta_{f,1}(\lambda_{1,h})_j+\ldots+\eta_{f,j}(\lambda_{j,h})_j=0\ (f\neq h)\text{ or }1\ (f=h). \qquad (41)$$

Let us write

$$\mathrm{H}_j\equiv\begin{vmatrix}\eta_{0,0} & \eta_{0,1} & \eta_{0,2} & \cdots & \eta_{0,j}\\ \eta_{1,0} & \eta_{1,1} & \eta_{1,2} & \cdots & \eta_{1,j}\\ \eta_{2,0} & \eta_{2,1} & \eta_{2,2} & \cdots & \eta_{2,j}\\ \cdot & \cdot & \cdot & & \cdot\\ \cdot & \cdot & \cdot & & \cdot\\ \cdot & \cdot & \cdot & & \cdot\\ \eta_{j,0} & \eta_{j,1} & \eta_{j,2} & \cdots & \eta_{j,j}\end{vmatrix}, \qquad (42)$$

$$\mathrm{H}_{g,h,j}\equiv\text{cofactor of }\eta_{g,h}\text{ in }\mathrm{H}_j=\mathrm{H}_{h,g,j}. \qquad (43)$$

Then

$$\eta_{f,0}\mathrm{H}_{0,h,j}+\eta_{f,1}\mathrm{H}_{1,h,j}+\ldots+\eta_{f,j}\mathrm{H}_{j,h,j}=0\ (f\neq h)\text{ or }\mathrm{H}_j\ (f=h)\,; \qquad (44)$$

and therefore, by (41),

$$(\lambda_{g,h})_j=\mathrm{H}_{g,h,j}/\mathrm{H}_j. \qquad (45)$$

Substituting in (40), we obtain $(\epsilon_h)_j$ in terms of the $\sigma$'s.

For the particular case of $g=h=j$, $(\lambda_{g,h})_j$ becomes $\Lambda_j$, and $\mathrm{H}_{g,h,j}$ becomes $\mathrm{H}_{j-1}$; so that

$$\Lambda_j=\mathrm{H}_{j-1}/\mathrm{H}_j. \qquad (46)$$

(iii.) As an example, suppose that we have several independent observations, of unequal accuracy, of a single quantity U, and that we wish to obtain a suitably weighted mean, which may be regarded as the improved value of any one of the observations. Let the observed values be $u_0$, $u_1$, $u_2$, ..., and their m.ss.e. $a_0^2$, $a_1^2$, $a_2^2$, ...; the m.pp.e. being 0, since the observations are independent. We take $\delta_0$ to be one of the $u$'s, and $\delta_1$, $\delta_2$, ... to be its successive differences. Then $j=0$, since the true values of the first and later differences are all 0. Hence, by § 8 (i.), the improved value is $\sigma_0$/(m.s.e. of $\sigma_0$). But, by § 5 (i.), $\sigma_0=\Sigma y$; and, by § 2 (vi.) ($b$), $y_r=u_r/a_r^2$, so that m.s.e. of $\sigma_0=\Sigma 1/a^2$. Hence the improved value is $\Sigma(u/a^2)/\Sigma(1/a^2)$.

(iv.) When $j$ is relatively large, the solution given in (ii.) above can only be regarded as a formal one, since it involves calculation of determinants. I have not been able to provide a general solution which shall avoid determinants; but it will be seen in §§ 17–19 that, if we can find the values of certain quantities occurring in the formulæ, we can deduce the $\lambda$'s and thence the coefficients of the $\sigma$'s. These latter are important as giving us formulæ which contain only a few terms and are therefore suited for numerical calculation.

10. *Expressions in terms of a Related Set.*—Suppose that there is another set of $l+1$ quantities $u_0, u_1, u_2, \ldots u_l$, connected with the $\delta$'s by linear relations; and let the set conjugate to the $u$'s be $y_0, y_1, y_2, \ldots y_l$. We shall take the relations between the $u$'s and the $\delta$'s and between the $y$'s and the $\delta$'s to be, as in § 4, ($r = 0, 1, 2, \ldots l$)

$$u_r = (r_0)\,\delta_0 + (r_1)\,\delta_1 + (r_2)\,\delta_2 + \ldots + (r_l)\,\delta_l, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (47)$$

$$y_r = [r_0]\,\delta_0 + [r_1]\,\delta_1 + [r_2]\,\delta_2 + \ldots + [r_l]\,\delta_l. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (48)$$

(i.) Let the improved value of $u_r$, using $\delta$'s after $\delta_j$, be $v_r$. Then, from (47), by (XIII.), remembering that, by (36),

$$(\epsilon_f)_j = 0 \text{ if } f > j,$$

we have

$$v_r = (r_0)\,(\epsilon_0)_j + (r_1)\,(\epsilon_1)_j + \ldots + (r_l)\,(\epsilon_l)_j \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (49)$$

$$= (r_0)\,(\epsilon_0)_j + (r_1)\,(\epsilon_1)_j + \ldots + (r_j)\,(\epsilon_j)_j. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (49\text{A})$$

Thus the $v$'s are related to the $\epsilon$'s in the same way that the $u$'s are related to the $\delta$'s.

(ii.) Similarly, if the improved value of $y_r$, using $\delta$'s after $\delta_j$, is $z_r$, we have

$$z_r = [r_0]\,(\epsilon_0)_j + [r_1]\,(\epsilon_1)_j + \ldots + [r_l]\,(\epsilon_l)_j \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (50)$$

$$= [r_0]\,(\epsilon_0)_j + [r_1]\,(\epsilon_1)_j + \ldots + [r_j]\,(\epsilon_j)_j\,; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (50\text{A})$$

and the $z$'s are related to the $\epsilon$'s in the same way that the $y$'s are related to the $\delta$'s.

(iii.) Let $w$ be any l.c. of the $\delta$'s or of the $u$'s or $y$'s, and let $x$ be its improved value, using $\delta$'s after $\delta_j$. Suppose that $x$ is expressed in terms of the $y$'s, the coefficients being $p_0, p_1, p_2, \ldots p_l$, so that

$$x = p_0 y_0 + p_1 y_1 + p_2 y_2 + \ldots + p_l y_l\,; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (51)$$

and let

$$\lambda_g \equiv \text{m.p.e. of } x \text{ and } (\epsilon_g)_j,$$

so that

$$\lambda_g = 0 \text{ if } g > j.$$

Then, by the condition of conjugacy of the $u$'s and the $y$'s,

$$\text{m.p.e. of } x \text{ and } u_r = p_r. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (52)$$

Hence, by (IV.) and (49) or (49A),

$$p_r = \text{m.p.e. of } x \text{ and } v_r$$

$$= (r_0)\lambda_0+(r_1)\lambda_1+(r_2)\lambda_2+\dots+(r_l)\lambda_l \quad . \quad . \quad . \quad . \quad . \quad . \quad (53)$$

$$= (r_0)\lambda_0+(r_1)\lambda_1+(r_2)\lambda_2+\dots+(r_j)\lambda_j . \quad . \quad . \quad . \quad . \quad . \quad . \quad (53\text{A})$$

Thus the $p$'s are related to the $\lambda$'s in the same way that the $v$'s are related to the $\epsilon$'s, or the $u$'s to the $\delta$'s.

(iv.) Similarly, if

$$x = q_0u_0+q_1u_1+q_2u_2+\dots+q_lu_l,$$

the $q$'s are related to the $\lambda$'s in the same way that the $z$'s are related to the $\epsilon$'s, or the $y$'s to the $\delta$'s.

11. *Special Case of Differences.*—The important practical case is that in which the $\delta$'s are successive differences of the $u$'s, in the general sense explained in §5 (iv.). If the differences of order exceeding $j$ are negligible, we can use them as auxiliaries for improving the $u$'s or the $\delta$'s or any l.c. of the $u$'s or the $\delta$'s.

(i.) Since the $\delta$'s are successive differences of the $u$'s, $(r_t)$ is (§5 (iv.)) a polynomial in $r$ of degree $t$.

(ii.) By §10 (i.) the $\epsilon$'s are the differences of the $v$'s according to the same system; and $v_r$ is a polynomial of degree $j$ in $r$, the differences of the $v$'s of order exceeding $j$ being zero.

(iii.) With the notation of §10 (iii.), the $\lambda$'s are the differences of the $p$'s according to the same system; and $p_r$ is a polynomial of degree $j$ in $r$, the differences of the $p$'s of order exceeding $j$ being zero.

(iv.) If we form the differences of the $u$'s in the usual way, there will be $l$ differences of order 1, $l-1$ of order 2, and so on. The $l-j+1$ of order $j$, namely $\Delta^j u_0$, $\Delta^j u_1$, ... $\Delta^j u_{l-j}$, will differ from one another by l.cc. of the differences of higher order; and therefore, by (XI.), they will have the same improved value. If we denote this by E, then, if $w \equiv p\Delta^j u_0+q\Delta^j u_1+r\Delta^j u_2+\dots$, the improved value of $w$ is $(p+q+r+\dots)$E.

## Relation of "Reduction of Error" to "Fitting" (of a Polynomial).

12. *Standard System.*—In the case of a standard system, the process of reduction of error and the process of fitting a polynomial (by least squares or by moments) give the same result. The following is a proof of this, not involving the properties of conjugate sets. The observed values are taken to be $u_0$, $u_1$, $u_2$, ... $u_l$; and $\sum_t$ denotes summation for $t = 0, 1, 2, \dots l$.

(i.) If the polynomial which we are fitting to the $u$'s is

$$v_q = a_0+a_1q+a_2q^2+\dots+a_jq^j, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

the values of the $a$'s when we fit by least squares are given ("Fitting," §§1, 2) by the equations ($f = 0, 1, 2, \ldots j$)

$$\sum_q q^f . a_0 + \sum_q q^{f+1} . a_1 + \ldots + \sum_q q^{f+j} . a_j = \sum_q q^f u_q \equiv \mathrm{M}_f . \quad . \quad . \quad . \quad . \quad (55)$$

These are the same equations that are given by the method of moments.

(ii.) The above equation (55) is a statement that the $f^{\text{th}}$ moment of the $v$'s is equal to that of the $u$'s. In order to prove that the process of reduction of error, using differences of order exceeding $j$ as auxiliaries, gives the same result, it is sufficient to show ($\alpha$) that the improved value of $u_q$ as given by this process is of the form of $v_q$ in (54), and ($\beta$) that the $f^{\text{th}}$ moment of the improved values of the $u$'s is equal to that of the original values for $f = 0, 1, 2, \ldots j$.

(iii.) We have shown, in §11 (ii.), that the improved value of $u_q$ is a polynomial of degree $j$ in $q$. This establishes ($\alpha$).

(iv.) By (XIII.), the $f^{\text{th}}$ moment of the improved values of the $u$'s is equal to the improved value of their $f^{\text{th}}$ moment. In order to show that this is equal to the original value of the $f^{\text{th}}$ moment, it is sufficient, by (IX.), to show that the m.p.e. of the original $f^{\text{th}}$ moment and every difference of order exceeding $j$ is zero.

(v.) Let the $k^{\text{th}}$ difference of $u_{r-k}$ be

$$\delta_k \equiv k_0 u_r - k_1 u_{r-1} + \ldots + (-)^k k_k u_{r-k}.$$

Then the $f^{\text{th}}$ moment is

$$\ldots + r^f u_r + (r-1)^f u_{r-1} + (r-2)^f u_{r-2} + \ldots ,$$

and the m.p.e. of this and $\delta_k$ is

$$k_0 r^f - k_1 (r-1)^f + k_2 (r-2)^f - \ldots .$$

But this is the $k^{\text{th}}$ difference of $(r-k)^f$, and is $= 0$ if $k > f$.

This proves the proposition.

13. *Fitting by Least Squares.*—Next suppose that the set is not self-conjugate. If the $\delta$'s were the differences of a set of $u$'s, we should fit a polynomial of degree (say) $j$ to the $u$'s. This suggests that, in the more general case, the $u$'s being connected with the $\delta$'s, as in § 10, by the relation ($r = 0, 1, 2, \ldots l$)

$$u_r = (r_0)\,\delta_0 + (r_1)\,\delta_1 + (r_2)\,\delta_2 + \ldots + (r_l)\,\delta_l, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (56)$$

we should try to fit an expression of the form

$$v_r = (r_0)\,\epsilon_0 + (r_1)\,\epsilon_1 + (r_2)\,\epsilon_2 + \ldots + (r_j)\,\epsilon_j \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (57)$$

to the $u$'s by an appropriate method of least squares; the $(r)$'s being the same as in (56), and the $\epsilon$'s being the quantities to be determined.

(i.) If the $y$'s are conjugate to the $u$'s, and if

$$\psi_{r,s} \equiv \text{m.p.e. of } y_r \text{ and } y_s = \psi_{s,r}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (58)$$

then (see Appendix II.) the direct (or *a priori*) probability of the occurrence of the given set of $u$'s, if the $v$'s as given by (57) were the true $U$'s, is proportional to

$$\exp\{-\tfrac{1}{2}\sum_r\sum_s \psi_{r,s}(u_r - v_r)(u_s - v_s)\},$$

where $\sum_t$ denotes summation for $t = 0, 1, 2, \ldots l$. The principle of the method of least squares therefore leads us* to choose the $\epsilon$'s so as to make

$$\sum_r\sum_s \psi_{r,s}(u_r - v_r)(u_s - v_s)$$

a minimum. Differentiating with regard to each of the $\epsilon$'s, this gives ($f = 0, 1, 2, \ldots j$)

$$\sum_s \{(0_f)\psi_{0,s} + (1_f)\psi_{1,s} + \ldots + (l_f)\psi_{l,s}\}(v_s - u_s) = 0. \quad . \quad . \quad . \quad . \quad (59)$$

But, by (58) and (16),

$$\begin{aligned}(0_f)\psi_{0,s} + (1_f)\psi_{1,s} + \ldots + (l_f)\psi_{l,s} &= \text{m.p.e. of } y_s \text{ and } (0_f)y_0 + (1_f)y_1 + \ldots + (l_f)y_l \\ &= \text{m.p.e. of } y_s \text{ and } \sigma_f. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (60)\end{aligned}$$

Denoting this, as in § 4 (iv.), by $[s_f]$, the equations (59) become ($f = 0, 1, 2, \ldots j$)

$$\sum_s [s_f](v_s - u_s) = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (61)$$

(ii.) Instead of fitting an expression of the form given by (57) to the $u$'s we might fit a corresponding expression to the $y$'s. Since

$$y_s = [s_0]\delta_0 + [s_1]\delta_1 + [s_2]\delta_2 + \ldots + [s_l]\delta_l, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (62)$$

the expression to be fitted would be of the form

$$z_s \equiv [s_0]\epsilon_0 + [s_1]\epsilon_1 + [s_2]\epsilon_2 + \ldots + [s_j]\epsilon_j. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (63)$$

---

[*] Strictly, we ought to choose the $\epsilon$'s so as to make $B \,.\, C \exp -\tfrac{1}{2}P$ a maximum: where

$$P \equiv \sum\sum \psi_{r,s}(u_r - v_r)(u_s - v_s);$$

$B$ is the direct probability of occurrence of the particular $\epsilon$'s denoted by $\epsilon_0, \epsilon_1, \epsilon_2, \ldots \epsilon_j$, and is therefore some function of these latter; and $C \exp -\tfrac{1}{2}P$ is the direct probability of occurrence of the particular values of $u - v$ on the assumption that these values of the $\epsilon$'s are the correct ones, $C$ being some function of these $\epsilon$'s. But I have assumed, as is commonly done, that the range of practically possible values of the $\epsilon$'s is so small that $B$ and $C$ may be treated as constants, so that we have only to consider the maximum value of $\exp -\tfrac{1}{2}P$.

We should have to choose the $\epsilon$'s so as to make

$$\sum_r \sum_s \pi_{r,s} (y_r - z_r)(y_s - z_s)$$

a minimum, where

$$\pi_{r,s} \equiv \text{m.p.e. of } u_r \text{ and } u_s. \qquad (64)$$

This would give

$$\sum_r (r_f)(z_r - y_r) = 0. \qquad (65)$$

(iii.) The $\epsilon$'s given by (65) are the same as are given by (61). For we have seen in § 4 that

$$(r_f) = \text{m.p.e. of } u_r \text{ and } \sigma_f, \qquad (66)$$

$$[s_f] = \text{m.p.e. of } y_s \text{ and } \sigma_f. \qquad (67)$$

If we express the $u$'s in terms of the $\delta$'s, and write

$$\sum_s [s_f] u_s = \sum_t A_t \delta_t,$$

then, by the condition of conjugacy of the $\sigma$'s and the $\delta$'s, and by (66),

$$\begin{aligned} A_t &= \text{m.p.e. of } \sigma_t \text{ and } \sum_s [s_f] u_s \\ &= \sum_s [s_f](s_t). \end{aligned}$$

This is symmetrical, and we should get the same expression for the coefficient of $\delta_t$ in $\sum_r (r_f) y_r$, so that

$$\sum_s [s_f] u_s = \sum_r (r_f) y_r. \qquad (68)$$

Similarly, if we substitute the values of $v_s$ from (57) and of $z_r$ from (63) in $\sum_s [s_f] v_s$ and in $\sum_r (r_f) z_r$, the coefficients of $\epsilon_t$ in the resulting expressions are equal. Hence (61) and (65) are identical equations in the $\epsilon$'s.

(iv.) The values of the $\epsilon$'s as given by these equations are in fact independent of the $u$'s or the $y$'s. For the value of $A_t$ as found in (iii.) above is

$$\begin{aligned} \sum_s [s_f](s_t) &= \text{m.p.e. of } \sum_s [s_f] u_s \text{ and } \sum_s (s_t) y_s \\ &= \text{m.p.e. of } \sigma_f \text{ and } \sigma_t, \qquad (69) \end{aligned}$$

by (24) and (16). Hence, denoting the m.p.e. of $\sigma_f$ and $\sigma_t$, as in § 3, by $\eta_{f,t}$, the $\epsilon$'s given by (61) or (65) are the same as would be given by ($f = 0, 1, 2, \ldots j$)

$$\sum_{t=0}^{t=j} \eta_{f,t} \epsilon_t = \sum_{t=0}^{t=l} \eta_{f,t} \delta_t. \qquad (70)$$

(v.) The ordinary method of least squares would consist of making $\sum_s (v_s - u_s)^2$ a minimum, and would lead to equations

$$\sum_s (s_f)(v_s - u_s) = 0,$$

which would not give the most probable values of the $\epsilon$'s.

14. *Fitting by Moments.*—(i.) The ordinary method of moments, adapted to the case in which the $\delta$'s are not necessarily the successive differences of the $u$'s, would consist in equating the values of $\sum_s (s_f) v_s$ and $\sum_s (s_f) u_s$. This, as will be seen from § 13 (v.), would not give the most probable values of the $\epsilon$'s.

(ii.) In order to obtain the most probable values of the $\epsilon$'s by equating moments of the $v$'s and of the $u$'s, we must write (say)

$$M_f \equiv \sum_s [s_f] u_s, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (71)$$

and define the $f^{\text{th}}$ moment of the $u$'s as being $M_f$ or a definite l.c. of $M_f$, $M_{f-1}$, $M_{f-2}, \ldots M_0$. But the coefficient of $u_s$ in $M_f$ would then not be given definitely by the relations between the $u$'s and the $\delta$'s, but would depend also on the law of correlation of errors of the $u$'s. We see, however, from § 13 (iii.), that we have also

$$M_f = \sum_r (r_f) y_r, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (72)$$

and that we get the same result by equating moments, defined in this way, of the $y$'s and the $z$'s. In the ordinary case in which the $\delta$'s are successive differences of the $u$'s, the coefficients of the $y$'s in (72) are binomial coefficients, and the ordinary moments fall within the definition given above. It follows that *in fitting a polynomial to a set of quantities* (not being a self-conjugate set) *by the method of moments, the moments which ought to be equated are not those of the quantities themselves and their assumed values, but those of the conjugate set of the former and the corresponding l.cc. of the latter.*

15. *Reduction of Error.*—If we improve the $\delta$'s or the $u$'s by means of the $\delta$'s after $\delta_j$, the improved values of these latter are zero, and those of the $\delta$'s up to $\delta_j$ are obtainable from (XXI.) of § 8, which states that the improved values of the $\sigma$'s from $\sigma_0$ to $\sigma_j$ are equal to the original values. Using (6), this gives ($f = 0, 1, 2, \ldots j$)

$$\sum_{t=0}^{t=j} \eta_{f,t} (\epsilon_t)_j = \sum_{t=0}^{t=l} \eta_{f,t} \delta_t. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (73)$$

Comparing this with (70), we see that the $\epsilon$'s given by this process are the same as those given by the process of fitting the expression in (57).

16. *Difference of the Two Processes.*—Although the two processes lead to the

same result, they are essentially different. This is explained in § 22 of "Reduction." The main difference may be expressed as follows:—

(i.) In "fitting" we deal directly with the particular case. We assume that the true values follow a specified law, involving unknown constants, and we deduce values for these constants from the data by the principle of inverse probability.

(ii.) In "reduction of error" we do not use inverse probability, and we only deal incidentally with the particular case. We regard the aggregate of the data as one of an indefinitely great number of possible aggregates from the same true values, and we use a method which will reduce as much as possible the m.s.e. of these possible aggregates.

## SOME STEPS IN THE GENERAL SOLUTION.

17. *Preliminary.*—(i.) Our object is to find the improved value of any l.c. of the $\delta$'s or the $u$'s, and the m.s.e. of this improved value or the m.p.e. of two improved values. Ordinarily, as already stated in § 8 (iv.), the quantity to be improved would be expressible in terms of the $\delta$'s, so that we need consider only the improved values of the $\delta$'s, *i.e.*, the $\epsilon$'s. There are then four problems before us, viz.: (1) expression of the $\epsilon$'s in terms of the $\delta$'s; (2) expression in terms of the $\sigma$'s; (3) expression in terms of the $y$'s; (4) determination of the $\lambda$'s. For practical purposes (2) is more important than (1) or (3), since there will be fewer coefficients involved.

(ii.) Although it does not seem possible to obtain a general solution, otherwise than by determinants, there are some general propositions that indicate stages in the solution. If, without necessarily finding the complete expressions of the $\epsilon$'s in terms of the $\delta$'s, we can find for each $\epsilon$ the coefficient of the first of the auxiliaries, then it will be seen from § 18 that we can find all the $\epsilon$'s if we know the E's, and from § 19 (i.) that we can find all the $\lambda$'s if we know the $\Lambda$'s. It follows from (40) that in this latter case we can at once obtain the $\epsilon$'s in terms of the $\sigma$'s.

(iii.) We use the notation of § 7, and we also write

$$-\theta_{f,j} \equiv \text{coefficient of } \delta_j \text{ (as auxiliary) in } (\epsilon_f)_{j-1},$$

so that

$$\theta_{f,j} = 0 \text{ if } f > j, \qquad (74)$$

$$\theta_{j,j} = 1. \qquad (75)$$

It should be observed that $\theta_{f,j}$ is not equal to $\theta_{j,f}$. The $\theta$'s may be known directly, or, as is shown in (83), we may be able to obtain them from certain of the $\lambda$'s.

18. *Formula of Progression.*—The quantities which we want to find are the improved values

of $\delta_0$, using $\delta_1, \delta_2, \delta_3, \ldots,$

of $\delta_1$ and $\delta_0$, using $\delta_2, \delta_3, \delta_4, \ldots,$

of $\delta_2$, $\delta_1$, and $\delta_0$, using $\delta_3, \delta_4, \delta_5, \ldots,$

and so on. There is a formula connecting these, which makes it unnecessary to deal with more than the first quantity in each row; or, if we deal with them all, the formula can be used for checking the results. (An example is given at the end of § 15 of "Reduction.")

We have

$$(\epsilon_f)_{j-1} = \delta_f - \theta_{f,j}\delta_j + \text{ terms in } \delta_{j+1},\ \delta_{j+2}, \ldots .$$

By (XV.), this is the improved value of $\delta_f - \theta_{f,j}\delta_j$, using $\delta$'s after $\delta_j$; and therefore, by (XIII.),

$$(\epsilon_f)_{j-1} = (\epsilon_f)_j - \theta_{f,j}(\epsilon_j)_j = (\epsilon_f)_j - \theta_{f,j}\mathrm{E}_j. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (76)$$

Re-arranging, and replacing $j$ by $j-1$, $j-2$, ... $f+1$, and remembering that, by (75), $\theta_{f,f} = 1$, we have

$$\begin{aligned}
(\epsilon_f)_j - (\epsilon_f)_{j-1} &= \theta_{f,j}\ \mathrm{E}_j, \\
(\epsilon_f)_{j-1} - (\epsilon_f)_{j-2} &= \theta_{f,j-1}\mathrm{E}_{j-1}, \\
&\ \vdots \\
(\epsilon_f)_{f+1} - (\epsilon_f)_f &= \theta_{f,f+1}\mathrm{E}_{f+1}, \\
(\epsilon_f)_f &= \theta_{f,f}\ \mathrm{E}_f.
\end{aligned}$$

Hence, by addition,

$$(\epsilon_f)_j = \theta_{f,f}\mathrm{E}_f + \theta_{f,f+1}\mathrm{E}_{f+1} + \theta_{f,f+2}\mathrm{E}_{f+2} + \ldots + \theta_{f,j}\mathrm{E}_j. \quad . \quad . \quad . \quad . \quad . \quad (77)$$

19. *Mean Products of Error (Alternative Formula).*—(i.) By (77) and (38),

$$\begin{aligned}
(\lambda_{f,g})_j &= \text{m.p.e. of } \delta_g \text{ and } (\epsilon_f)_j \\
&= \text{m.p.e. of } \delta_g \text{ and } \theta_{f,f}\mathrm{E}_f + \theta_{f,f+1}\mathrm{E}_{f+1} + \ldots + \theta_{f,j}\mathrm{E}_j \\
&= \theta_{f,f}(\lambda_{g,f})_f + \theta_{f,f+1}(\lambda_{g,f+1})_{f+1} + \ldots + \theta_{f,j}(\lambda_{g,j})_j \\
&= \Sigma\theta_{f,t}(\lambda_{g,t})_t.
\end{aligned}$$

The summation has to be made from $t = f$ to $t = j$. But, if $g > f$, we see from (37) that it is sufficient to make the summation from $t = g$. Hence, using "$t = f, g$" to denote summation from $t = f$ or from $t = g$, according as $f$ or $g$ is the greater, we have

$$(\lambda_{f,g})_j = \sum_{t=f,g}^{t=j} \theta_{f,t}(\lambda_{g,t})_t. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (78)$$

But, by putting $g = j$ in (78) (or taking the m.p.e. of $\delta_j$ and each member of (76)) and then replacing $f$ and $j$ by $g$ and $t$,

$$(\lambda_{g,t})_t = \theta_{g,t}\Lambda_t. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (79)$$

Hence, substituting in (78),

$$(\lambda_{f,g})_j = \sum_{t=f,g}^{t=j} \theta_{f,t}\theta_{g,t}\Lambda_t. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (80)$$

If we can obtain the $\theta$'s and the $\Lambda$'s in a simple form, we thus have a workable formula for calculating the $\lambda$'s, and thence, by (40), for determining the $\epsilon$'s in terms of the $\sigma$'s.

(ii.) From (80), using (II.) of § 6, we get the m.ss.e. and m.pp.e. of the improved values of any l.cc. of the $\delta$'s. Let

$$w \equiv b_0\delta_0+b_1\delta_1+\ldots+b_l\delta_l, \qquad w' \equiv c_0\delta_0+c_1\delta_1+\ldots+c_l\delta_l,$$

and let the improved values of $w$ and $w'$, using $\delta$'s after $\delta_j$, be $x$ and $x'$. Then

$$\begin{aligned}\text{m.p.e. of } x \text{ and } x' &= \sum_{t=0}^{t=j}(b_0\theta_{0,t}+b_1\theta_{1,t}+\ldots)(c_0\theta_{0,t}+c_1\theta_{1,t}+\ldots)\Lambda_t\\ &= \sum_{t=0}^{t=j}(b_0\theta_{0,t}+b_1\theta_{1,t}+\ldots+b_t\theta_{t,t})(c_0\theta_{0,t}+c_1\theta_{1,t}+\ldots+c_t\theta_{t,t})\Lambda_t, \quad (81)\end{aligned}$$

$$\text{m.s.e. of } x = \sum_{t=0}^{t=j}(b_0\theta_{0,t}+b_1\theta_{1,t}+\ldots+b_t\theta_{t,t})^2\Lambda_t. \quad . \quad . \quad . \quad . \quad . \quad (82)$$

(iii.) We have assumed that the $\theta$'s are known. If they are not known directly, but the values of $(\lambda_{f,t})_t$ are known, then, by (79),

$$\theta_{f,t} = (\lambda_{f,t})_t/\Lambda_t. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (83)$$

Substituting in (80),

$$(\lambda_{f,g})_j = \sum_{t=f,g}^{t=j}(\lambda_{f,t})_t(\lambda_{g,t})_t/\Lambda_t. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (84)$$

Also (77) is replaced by

$$(\epsilon_f)_j = \sum_{t=f}^{t=j}(\lambda_{f,t})_t\mathrm{E}_t/\Lambda_t. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (85)$$

## APPLICATION TO SELF-CONJUGATE SET.

20. *Preliminary.*—(i.) We have now to apply the preceding results to the case in which the $u$'s are a self-conjugate set, so that $(u_r\,;\,u_s) = 0$ $(r\neq s)$, $(u_r\,;\,u_r) = 1$, $y_r = u_r$. We take the $\delta$'s to be successive differences of the $u$'s, commencing with a difference of order 0. The $\delta$'s to be used as auxiliaries will be those following $\delta_j$; the $(\quad)_j$ will usually be omitted. We shall take the number of $u$'s or of $\delta$'s to be $m$, so that $m = l+1$.

(ii.) There will be three cases to be considered; advancing differences, and central differences for $m$ odd and for $m$ even. For advancing differences the $u$'s will be taken to be $u_0, u_1, \ldots u_{m-1}$. For central differences we shall write $m = 2n+1$ or $m = 2n$; and the $u$'s will be $u_{-n}, u_{-n+1}, \ldots u_n$ and $u_{-n+1}, u_{-n+2}, \ldots u_n$ respectively. We shall require the following m.pp.e., which can be obtained from ordinary difference formulæ.

$$\text{m.p.e. of } \Delta^f u_0 \text{ and } \Delta^g u_0 = (-)^{f-g}(f+g, f), \quad \ldots \quad (86)$$

$$\text{,,} \quad \delta^{2f} u_0 \quad \text{,,} \quad \delta^{2g} u_0 = (-)^{f-g}(2f+2g, f+g), \quad \ldots \quad (87)$$

$$\text{,,} \quad \delta^{2f} u_0 \quad \text{,,} \quad \mu\delta^{2g-1} u_0 = 0, \quad \ldots \quad (88)$$

$$\text{,,} \quad \mu\delta^{2f-1} u_0 \quad \text{,,} \quad \mu\delta^{2g-1} u_0 = (-)^{f-g}(2f+2g-2, f+g-1)/(2f+2g), \quad . \quad (89)$$

$$\text{,,} \quad \delta^{2f-1} u_{\frac{1}{2}} \quad \text{,,} \quad \delta^{2g-1} u_{\frac{1}{2}} = (-)^{f-g}(2f+2g-2, f+g-1), \quad \ldots \quad (90)$$

$$\text{,,} \quad \delta^{2f-1} u_{\frac{1}{2}} \quad \text{,,} \quad \mu\delta^{2g-2} u_{\frac{1}{2}} = 0, \quad \ldots \quad (91)$$

$$\text{,,} \quad \mu\delta^{2f-2} u_{\frac{1}{2}} \quad \text{,,} \quad \mu\delta^{2g-2} u_{\frac{1}{2}} = (-)^{f-g}(2f+2g-4, f+g-2)/(2f+2g-2). \quad (92)$$

(iii.) For advancing differences we shall have

$$\delta^f \equiv \Delta^f u_0, \qquad \epsilon_f \equiv \Delta^f v_0.$$

The formulæ will be marked (A).

(iv.) For central differences the two cases of $m$ odd and $m$ even must be considered separately; but it will be found that, when the formulæ relating to $v_0$, $\delta^2 v_0$, ... ($m$ odd) and to $\delta v_{\frac{1}{2}}$, $\delta^3 v_{\frac{1}{2}}$, ... ($m$ even) are properly expressed, they are practically identical in form, as also are those relating to $\mu\delta v_0$, $\mu\delta^3 v_0$ ... ($m$ odd) and to $\mu v_{\frac{1}{2}}$, $\mu\delta^2 v_{\frac{1}{2}}$, ... ($m$ even); and the latter correspond to the former with certain interchanges of ( ] and [ ). We therefore, for $\mu\delta v_0$, $\mu\delta^3 v_0$, ... and $\mu v_{\frac{1}{2}}$, $\mu\delta^2 v_{\frac{1}{2}}$, ..., replace $\theta$, E, $\Lambda$, $\lambda$, by $\phi$, I, M, $\mu$, with the appropriate suffixes.

(v.) For $m = 2n+1$ it will be seen from (88), taken with (XIV.) of § 6, that the differences of even and of odd order can be treated independently. The $\delta$'s will be $u_0$, $\delta^2 u_0$, ... $\delta^{2n} u_0$ in the one case and $\mu\delta u_0$, $\mu\delta^3 u_0$, ... $\mu\delta^{2n-1} u_0$ in the other. We shall denote these by $\delta_0$, $\delta_2$, ... $\delta_{2n}$ and $\delta_1$, $\delta_3$, ... $\delta_{2n-1}$ respectively, and shall take $j$ to be $2k$ or $2k+1$ for the former and $2k-1$ or $2k$ for the latter. The subscripts of the $\theta$'s and the $\phi$'s will be modified accordingly; *i.e.*, $\theta_{2f, 2k}$ will mean the coefficient of $-\delta_{2k}$ in $(\epsilon_{2f})_{2k-2}$, and similarly for $\phi_{2f-1, 2k-1}$. The formulæ for the two cases will be marked (B) and (C) respectively.

(vi.) Similarly for $m = 2n$ we see from (91) that differences of odd and of even order can be treated separately. The $\delta$'s are $\delta_1$, $\delta_3$, ... $\delta_{2n-1}$ in the one case, and $\delta_0$, $\delta_2$, ... $\delta_{2n-2}$ in the other, where $\delta_{2f-1} \equiv \delta^{2f-1} u_{\frac{1}{2}}$, $\delta_{2f-2} \equiv \mu\delta^{2f-2} u_{\frac{1}{2}}$; and $j$ is taken to be $2k-1$ or $2k$ for the former and $2k-2$ or $2k-1$ for the latter. Also $\theta_{2f-1, 2k-1}$ means the coefficient of $-\delta_{2k-1}$ in $(\epsilon_{2f-1})_{2k-3}$; and similarly for $\phi_{2f-2, 2k-2}$. The formulæ will be marked (D) and (E).

(vii.) Writing

$$F\{\alpha, \beta, \ldots ; \rho, \psi, \ldots\} \equiv 1 + \frac{\alpha . \beta \ldots}{\rho . \psi \ldots} + \frac{\alpha(\alpha+1) . \beta(\beta+1) \ldots}{\rho(\rho+1) . \psi(\psi+1) \ldots} + \ldots,$$

where $\alpha$ is a negative integer, it should be noted that

$$F\{-n, \beta ; 1, \psi\} = \frac{[\psi-\beta, n]}{[\psi, n]}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (93)$$

and that, if

$$\alpha = -n, \quad \psi+\chi = -n+\beta+\gamma+1,$$

then*

$$F\{-n, \beta, \gamma ; 1, \psi, \chi\} = \frac{[\psi-\beta, n][\chi-\beta, n]}{[\psi, n][\chi, n]} = \frac{[\psi-\gamma, n][\chi-\gamma, n]}{[\psi, n][\chi, n]}. \quad . \quad . \quad (94)$$

(viii.) For the central-difference formulæ it will be convenient to write, if $r$ and $s$ are both even or both odd and $s \geqslant r$,

$$\{s, r\} \equiv (\tfrac{1}{2})^s \frac{(s+r, s)(s, \frac{1}{2}s+\frac{1}{2}r)}{r+1}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (95)$$

$$\{r, s\} \equiv (\tfrac{1}{2})^r \frac{(s+r, r)}{(r+1)(\frac{1}{2}r+\frac{1}{2}s, r)} ; \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (96)$$

so that, if $k \geqslant f$,

$$\{2k+1, 2f+1\} = \frac{[f+\frac{3}{2}, k](k, f)}{2f+2} = \frac{[f+\frac{1}{2}, k+1](k+1, f+1)}{2f+1}, \quad . \quad . \quad (97)$$

$$\{2k, 2f\} = \frac{[f+\frac{1}{2}, k](k, f)}{2f+1} = \frac{[f+\frac{3}{2}, k-1](k-1, f-1)}{2f}, \quad . \quad . \quad (98)$$

and, if $k \leqslant s$,

$$\{2k+1, 2s+1\} = \frac{[s+\frac{3}{2}, k]}{(2k+2)(s, k)} = \frac{[s+\frac{1}{2}, k+1]}{(2s+1)(s, k)}, \quad . \quad . \quad . \quad . \quad (99)$$

$$\{2k, 2s\} = \frac{[s+\frac{1}{2}, k]}{(2k+1)(s, k)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (100)$$

(ix.) The successive steps are as follows. The formulæ for the $\epsilon$'s (the improved values of the differences) in terms of the differences have already been found in "Reduction" and "Fitting"; they depend on certain theorems as to the coefficients when l.cc. of moments or sums are expressed in terms of differences. From these formulæ we get the $\theta$'s, and also the E's; and thence we get, in each case, the progression formula supplied by (77). This formula is not really necessary, but it is useful for checking. The $\Lambda$'s, *i.e.* the m.ss.e. of the E's, are found from (86)–(92), using (IV.) of § 6; this results in certain hypergeometric series, to which we apply (93) and (94). We then get expressions for the $\lambda$'s, by (80). From these, by (40), we have the values of the $\epsilon$'s in terms of the $\sigma$'s; and also, by § 10 (iii.), the values in terms of the $u$'s (which are identical with the $y$'s). This completes the investigation of the improved values; but we also want to see the extent of the improvement. A further section therefore gives the ratio of the m.s.e. of the improved value to the m.s.e. of the original value, in a form convenient for calculation.

---

* 'Proceedings of the London Mathematical Society,' 2nd series, x., 474.

(x.) The m.p.e. of $\Delta^f u_r$ and $\Delta^f u_{r+t}$ is $(-)^t\,(2f, f+t)$. Hence, in finding the general solution of our problem for the case of a self-conjugate set, we are also finding it for the case of a set in which the m.p.e. of $u_r$ and $u_s$ is of the form $(-)^{sr}\,(2f, f+s-r)$, $f$ being some positive integer; for we can treat these $u$'s as the $f^{\text{th}}$ differences of members of a self-conjugate set.

21. *Improved Values.*—(i.) Adapting § 11 (iii.) to the case in which the $y$'s are identical with the $u$'s, we see that, if $w$ is any l.c. of the $u$'s, its improved value, using differences of order exceeding $j$, is of the form $\Sigma p_r u_r$, where $p_r$ is a polynomial of degree $j$ in $r$. The improved values of the $u$'s and their differences are obtained from this by means of certain formulæ, given in §§ 6 and 7 (iv.) of "Factorial Moments," for the expression of $\Sigma p_r u_r$ in terms of differences. The results are given in (12), (18), (19), (26), and (25) of "Fitting." From the first of these, replacing $m+1$ by $m$, we have ($f = 0, 1, 2, \ldots j$)

$$\text{(A)}\qquad \Delta^f v_0 = \sum_{s=0}^{s=m-1} (s, f)\,(j-s, j-f)\,\frac{(j+f+1, j)}{(j+s+1, j)}\,\frac{(m, s+1)}{(m, f+1)}\,\Delta^s u_0\,; \qquad (101)$$

and from the other four, altering $k$ to $k-1$ in the last, we have ($f = 0, 1, 2, \ldots k$ in (102) and $1, 2, \ldots k$ in (103)-(105))

$$\text{(B)}\qquad \delta^{2f} v_0 = \sum_{s=0}^{s=n} (k+s, k+f)\,(k-s, k-f)\,\frac{(2k+2f+1, 2k)}{(2k+2s+1, 2k)}\,\frac{[n+\tfrac{1}{2}, 2s+1)}{[n+\tfrac{1}{2}, 2f+1)}\,\delta^{2s} u_0\,. \qquad (102)$$

$$= \delta^{2f} u_0 + (-)^{k-f} \sum_{s=k+1}^{s=n} \frac{s-k}{s-f}\,\frac{2f+2}{2k+2}\,\frac{\{2k+1, 2f+1\}}{\{2k+1, 2s+1\}}\,\frac{[\tfrac{1}{2}m, 2s+1)}{[\tfrac{1}{2}m, 2f+1)}\,\delta^{2s} u_0, \qquad (102\text{A})$$

$$\text{(C)}\quad \mu\delta^{2f-1} v_0 = \sum_{s=1}^{s=n} (k+s-1, k+f-1)\,(k-s, k-f)\,\frac{(2k+2f-1, 2k-1)}{(2k+2s-1, 2k-1)}\,\frac{(n+\tfrac{1}{2}, 2s]}{(n+\tfrac{1}{2}, 2f]}\,\mu\delta^{2s-1} u_0 \qquad (103)$$

$$= \mu\delta^{2f-1} u_0 + (-)^{k-f} \sum_{s=k+1}^{s=n} \frac{s-k}{s-f}\,\frac{2f+1}{2k+1}\,\frac{\{2k, 2f\}}{\{2k, 2s\}}\,\frac{(\tfrac{1}{2}m, 2s]}{(\tfrac{1}{2}m, 2f]}\,\mu\delta^{2s-1} u_0, \qquad (103\text{A})$$

$$\text{(D)}\qquad \delta^{2f-1} v_{\frac{1}{2}} = \sum_{s=1}^{s=n} (k+s-1, k+f-1)\,(k-s, k-f)\,\frac{(2k+2f-1, 2k-1)}{(2k+2s-1, 2k-1)}\,\frac{[n, 2s)}{[n, 2f)}\,\delta^{2s-1} u_{\frac{1}{2}} \qquad (104)$$

$$= \delta^{2f-1} u_{\frac{1}{2}} + (-)^{k-f} \sum_{s=k+1}^{s=n} \frac{s-k}{s-f}\,\frac{2f+1}{2k+1}\,\frac{\{2k, 2f\}}{\{2k, 2s\}}\,\frac{[\tfrac{1}{2}m, 2s)}{[\tfrac{1}{2}m, 2f)}\,\delta^{2s-1} u_{\frac{1}{2}}, \qquad (104\text{A})$$

(E) $$\mu\delta^{2f-2}v_{\frac{1}{2}} = \sum_{s=1}^{s=n}(k+s-2,\ k+f-2)(k-s,\ k-f)\frac{(2k+2f-3,\ 2k-2)}{(2k+2s-3,\ 2k-2)}\frac{(n,\ 2s-1]}{(n,\ 2f-1]}\mu\delta^{2s-2}u_{\frac{1}{2}} \quad . \ . \ . \ (105)$$

$$= \mu\delta^{2f-2}u_{\frac{1}{2}} + (-)^{k-f}\sum_{s=k+1}^{s=n}\frac{s-k}{s-f}\frac{2f}{2k}\frac{\{2k-1,\ 2f-1\}}{\{2k-1,\ 2s-1\}}\frac{(\frac{1}{2}m,\ 2s-1]}{(\frac{1}{2}m,\ 2f-1]}\mu\delta^{2s-2}u_{\frac{1}{2}}. \quad . \ . \ . \ (105\text{A})$$

(ii.) From (i.) we obtain

(A) $$\theta_{f,j} = (-)^{j-f}(j,f)\frac{(j+f,\ j-1)}{(2j,\ j-1)}\frac{(m,\ j+1)}{(m,\ f+1)}, \quad . \ . \ . \ . \ (106)$$

(B) $$\theta_{2f,2k} = (-)^{k-f}\frac{\{2k,\ 2f\}}{\{2k,\ 2k\}}\frac{[\frac{1}{2}m,\ 2k+1)}{[\frac{1}{2}m,\ 2f+1)}, \quad . \ . \ . \ . \ . \ . \ . \ (107)$$

(C) $$\phi_{2f-1,2k-1} = (-)^{k-f}\frac{\{2k-1,\ 2f-1\}}{\{2k-1,\ 2k-1\}}\frac{(\frac{1}{2}m,\ 2k]}{(\frac{1}{2}m,\ 2f]}, \quad . \ . \ . \ . \ . \ (108)$$

(D) $$\theta_{2f-1,2k-1} = (-)^{k-f}\frac{\{2k-1,\ 2f-1\}}{\{2k-1,\ 2k-1\}}\frac{[\frac{1}{2}m,\ 2k)}{[\frac{1}{2}m,\ 2f)}, \quad . \ . \ . \ . \ . \ (109)$$

(E) $$\phi_{2f-2,2k-2} = (-)^{k-f}\frac{\{2k-2,\ 2f-2\}}{\{2k-2,\ 2k-2\}}\frac{(\frac{1}{2}m,\ 2k-1]}{(\frac{1}{2}m,\ 2f-1]}. \quad . \ . \ . \ . \ . \ (110)$$

(iii.) Also, by putting $f=j$ in (101) and $f=k$ in (102)–(105),

(A) $$\mathrm{E}_j = \Delta^j v_0 = \sum_{s=j}^{s=m-1}(s,\ j)\frac{(2j+1,\ j)}{(j+s+1,\ j)}\frac{(m,\ s+1)}{(m,\ j+1)}\Delta^s u_0, \quad . \ . \ . \ (111)$$

(B) $$\mathrm{E}_{2k} = \delta^{2k}v_0 = \sum_{s=k}^{s=n}\frac{\{2k+1,\ 2k+1\}}{\{2k+1,\ 2s+1\}}\frac{[\frac{1}{2}m,\ 2s+1)}{[\frac{1}{2}m,\ 2k+1)}\delta^{2s}u_0, \quad . \ . \ . \ . \ (112)$$

(C) $$\mathrm{I}_{2k-1} = \mu\delta^{2k-1}v_0 = \sum_{s=k}^{s=n}\frac{\{2k,\ 2k\}}{\{2k,\ 2s\}}\frac{(\frac{1}{2}m,\ 2s]}{(\frac{1}{2}m,\ 2k]}\mu\delta^{2s-1}u_0, \quad . \ . \ . \ . \ . \ . \ (113)$$

(D) $$\mathrm{E}_{2k-1} = \delta^{2k-1}v_{\frac{1}{2}} = \sum_{s=k}^{s=n}\frac{\{2k,\ 2k\}}{\{2k,\ 2s\}}\frac{[\frac{1}{2}m,\ 2s)}{[\frac{1}{2}m,\ 2k)}\delta^{2s-1}u_{\frac{1}{2}}, \quad . \ . \ . \ . \ . \ . \ (114)$$

(E) $$\mathrm{I}_{2k-2} = \mu\delta^{2k-2}v_{\frac{1}{2}} = \sum_{s=k}^{s=n}\frac{\{2k-1,\ 2k-1\}}{\{2k-1,\ 2s-1\}}\frac{(\frac{1}{2}m,\ 2s-1]}{(\frac{1}{2}m,\ 2k-1]}\mu\delta^{2s-2}u_{\frac{1}{2}}. \quad . \ . \ . \ (115)$$

(iv.) It has been pointed out in § 11 (iv.) that the differences of order $j$ all have the same improved value. It follows that (112)–(115) are particular cases of (111), expressed in terms of central differences; the proper values being taken for $m$ and for $j$, and $u_0$ being altered to $u_{-n}$ for (112) and (113) and to $u_{-n+1}$ for (114) and (115). We can verify this by expressing the E's in terms of the differences of order $j$. For (A) we have

$$\Delta^s u_0 = \{(1+\Delta)-1\}^{s-j}\Delta^j u_0 = \Delta^j u_{s-j} - (s-j,\ 1)\,\Delta^j u_{s-j-1} + \dots .$$

Substituting in (111), and rearranging the terms, it will be found that the coefficient of $\Delta^j u_r$ is

$$(j+r,\, j)\frac{(2j+1,\, j)\,(m,\, j+r+1)}{(2j+r+1,\, j)\,(m,\, j+1)}\,F\{-m+j+r+1,\, j+r+1\ ;\ 1,\, 2j+r+2\}$$
$$= (j+r,\, j)\,(m-r-1,\, j)/(m,\, 2j+1]\,;$$

so that

$$\text{(A)}\qquad \mathrm{E}_j = \sum_{r=0}^{r=m-j-1}(j+r,\, j)\frac{(m-r-1,\, j)}{(m,\, 2j+1]}\,\Delta^j u_r. \qquad \ldots\ldots \qquad (116)$$

Similarly from (112)–(115)

$$\text{(B)}\qquad \mathrm{E}_{2k} = \sum_{r=-n+k}^{r=n-k}\frac{(n+k+r,\, 2k)\,(n+k-r,\, 2k)}{(m,\, 4k+1]}\,\delta^{2k}u_r, \qquad \ldots\ldots \qquad (117)$$

$$\text{(C)}\qquad \mathrm{I}_{2k-1} = \sum_{r=0}^{r=n-k}\frac{(n+k+r,\, 2k-1)\,(n+k-r-1,\, 2k-1)}{(m,\, 4k-1]}\,(\delta^{2k-1}u_{-r-\frac{1}{2}}+\delta^{2k-1}u_{r+\frac{1}{2}}), \quad . \quad (118)$$

$$\text{(D)}\qquad \mathrm{E}_{2k-1} = \sum_{r=-n+k}^{r=n-k}\frac{(n+k+r-1,\, 2k-1)\,(n+k-r-1,\, 2k-1)}{(m,\, 4k-1]}\,\delta^{2k-1}u_{r+\frac{1}{2}}, \qquad \ldots \qquad (119)$$

$$\text{(E)}\qquad \mathrm{I}_{2k-2} = \sum_{r=0}^{r=n-k}\frac{(n+k+r-1,\, 2k-2)\,(n+k-r-2,\, 2k-2)}{(m,\, 4k-3]}\,(\delta^{2k-2}u_{-r}+\delta^{2k-2}u_{r+1}). \quad . \quad (120)$$

The identity of (117)–(120) with (116), the $u$'s being altered as explained above, is easily verified.

(v.) The formula of progression (77) takes simple forms if we attach the factors involving $m$ to the $\delta$'s; for $m$ then disappears from the formula.

(A) Writing

$$A_t \equiv (m,\, t+1)\,\mathrm{E}_t = (m,\, t+1)\,(\Delta^t v_0)_t,$$

so that, in effect, we take $\delta_f$ to be $(m,\, f+1)\,\Delta^f u_0$, (77) gives

$$\text{(A)}\qquad (m,\, f+1)\,\Delta^f v_0 = \sum_{t=f}^{t=j}(-)^{t-f}(t,\, f)\,\frac{(t+f,\, t-1)}{(2t,\, t-1)}\,A_t. \qquad \ldots\ldots \qquad (121)$$

For $j = 3$, for instance, we should have

$$\left.\begin{aligned}
(m,\, 1)\,v_0 &= A_0 - A_1 + \tfrac{1}{2}A_2 - \tfrac{1}{5}A_3 \\
(m,\, 2)\,\Delta v_0 &= \phantom{A_0 - {}} A_1 - \tfrac{3}{2}A_2 + \tfrac{6}{5}A_3 \\
(m,\, 3)\,\Delta^2 v_0 &= \phantom{A_0 - A_1 + \tfrac{1}{2}} A_2 - 2A_3 \\
(m,\, 4)\,\Delta^3 v_0 &= \phantom{A_0 - A_1 + \tfrac{1}{2}A_2 - \tfrac{1}{5}} A_3
\end{aligned}\right\},$$

where $A_0$ is the value of $(m,\, 1)\,v_0$ for $j = 0$, $A_1$ is the value of $(m,\, 2)\,\Delta v_0$ for $j = 1$, and so on. This may be verified by § 5 (i.)–(iii.) of "Fitting."

(B) (C) Writing

$$P_{2t} \equiv [\tfrac{1}{2}m,\, 2t+1)\,\mathrm{E}_{2t},\quad Q_{2t-1} \equiv (\tfrac{1}{2}m,\, 2t]\,\mathrm{I}_{2t-1},$$

we have

$$\text{(B)} \qquad [\tfrac{1}{2}m, 2f+1)\,\delta^{2f}v_0 = \sum_{t=f}^{t=k}(-)^{t-f}\frac{\{2t, 2f\}}{\{2t, 2t\}}P_{2t}, \qquad (122)$$

$$\text{(C)} \qquad (\tfrac{1}{2}m, 2f]\,\mu\delta^{2f-1}v_0 = \sum_{t=f}^{t=k}(-)^{t-f}\frac{\{2t-1, 2f-1\}}{\{2t-1, 2t-1\}}Q_{2t-1}. \qquad (123)$$

The first of these has been given, for $f = 0$, in "Reduction," § 15 (v.), p. 362; the notation of $P_{2t}$ differing, however, in a factor $\frac{1}{2}m$.

(D) (E) Writing

$$P_{2t-1} \equiv [\tfrac{1}{2}m, 2t)\,\mathrm{E}_{2t-1}, \quad Q_{2t-2} \equiv (\tfrac{1}{2}m, 2t-1]\,\mathrm{I}_{2t-2},$$

we have

$$\text{(D)} \qquad [\tfrac{1}{2}m, 2f)\,\delta^{2f-1}v_{\frac{1}{2}} = \sum_{t=f}^{t=k}(-)^{t-f}\frac{\{2t-1, 2f-1\}}{\{2t-1, 2t-1\}}P_{2t-1}, \qquad (124)$$

$$\text{(E)} \qquad (\tfrac{1}{2}m, 2f-1]\,\mu\delta^{2f-2}v_{\frac{1}{2}} = \sum_{t=f}^{t=k}(-)^{t-f}\frac{\{2t-2, 2f-2\}}{\{2t-2, 2t-2\}}Q_{2t-2}. \qquad (125)$$

(vi.) The Λ's are obtained from the E's by means of (IV.) of § 6; *e.g.*, for advancing differences, m.s.e. of $\mathrm{E}_j$ = m.p.e. of $\mathrm{E}_j$ and $\Delta^j u_0$. We can use either the values of the E's given by (111)–(115) or those given by (116)–(120); for the former we require the m.pp.e. given in (86), (87), (89), (90), (92), and for the latter the value of the m.p.e. of two differences of order $f$ as given in § 20 (x.). Using the former method, we get the following results:—

$$\begin{aligned}
\text{(A)} \qquad \Lambda_j &= \sum_{s=j}^{s=m-1}(-)^{s-j}(s, j)\frac{(2j+1, j)}{(s+j+1, j)}\frac{(m, s+1)}{(m, j+1)}(s+j, j) \\
&= (2j, j)\,F\{-m+j+1, 2j+1\,;\ 1, 2j+2\} \\
&= (2j, j)\,(m-j-1)\,!\,(2j+1)\,!/(m+j)\,! \\
&= (2j, j)/(m, 2j+1]\,; \qquad (126)
\end{aligned}$$

$$\begin{aligned}
\text{(B)} \qquad \Lambda_{2k} &= \sum_{s=k}^{s=n}(-)^{s-k}\frac{\{2k+1, 2k+1\}}{\{2k+1, 2s+1\}}\frac{[\frac{1}{2}m, 2s+1)}{[\frac{1}{2}m, 2k+1)}(2s+2k, s+k) \\
&= (4k, 2k)\,F\{-n+k, n+k+1, 2k+\tfrac{1}{2}\,;\ 1, 2k+1, 2k+\tfrac{3}{2}\} \\
&= (4k, 2k)/(m, 4k+1]\,; \qquad (127)
\end{aligned}$$

$$\begin{aligned}
\text{(C)} \qquad \mathrm{M}_{2k-1} &= \sum_{s=k}^{s=n}(-)^{s-k}\frac{\{2k, 2k\}}{\{2k, 2s\}}\frac{(\frac{1}{2}m, 2s]}{(\frac{1}{2}m, 2k]}\frac{(2s+2k-2, s+k-1)}{2s+2k} \\
&= (4k-2, 2k-1)/(4k)\,.\,F\{-n+k, n+k+1, 2k-\tfrac{1}{2}\,;\ 1, 2k+1, 2k+\tfrac{1}{2}\} \\
&= (4k-2, 2k-1)/(m, 4k-1]\,; \qquad (128)
\end{aligned}$$

$$\begin{aligned}
\text{(D)} \qquad \Lambda_{2k-1} &= \sum_{s=k}^{s=n}(-)^{s-k}\frac{\{2k, 2k\}}{\{2k, 2s\}}\frac{[\frac{1}{2}m, 2s)}{[\frac{1}{2}m, 2k)}(2s+2k-2, s+k-1) \\
&= (4k-2, 2k-1)\,F\{-n+k, n+k, 2k-\tfrac{1}{2}\,;\ 1, 2k, 2k+\tfrac{1}{2}\} \\
&= (4k-2, 2k-1)/(m, 4k-1]\,; \qquad (129)
\end{aligned}$$

$$\text{(E)}\qquad \mathrm{M}_{2k-2} = \sum_{s=k}^{s=n} (-)^{s-k} \frac{\{2k-1,\, 2k-1\}}{\{2k-1,\, 2s-1\}} \frac{(\tfrac{1}{2}m,\, 2s-1]}{(\tfrac{1}{2}m,\, 2k-1]} \frac{(2s+2k-4,\, s+k-2)}{2s+2k-2}$$

$$= (4k-4,\, 2k-2)/(4k-2) \,.\, F\{-n+k,\, n+k,\, 2k-\tfrac{3}{2}\,;\, 1,\, 2k,\, 2k-\tfrac{1}{2}\}$$

$$= (4k-4,\, 2k-2)/(m,\, 4k-3]. \qquad (130)$$

It would, of course, in view of the identity of the E's as stated in (iv.) above, have been sufficient to obtain (126) and deduce (127)–(130) from it.

(vii.) The Λ's having been found, the λ's, *i.e.*, the m.pp.e. of the improved values of $\Delta^f u_0$ and $\Delta^g u_0$, etc., are obtained by (80). The θ's and φ's being as in (ii.) above,

$$\text{(A)}\qquad \lambda_{f,g} = \sum_{t=f,g}^{t=j} \theta_{f,t}\,\theta_{g,t}\,(2t,\, t)/(m,\, 2t+1], \qquad (131)$$

$$\text{(B)}\qquad \lambda_{2f,2g} = \sum_{t=f,g}^{t=k} \theta_{2f,2t}\,\theta_{2g,2t}\,(4t,\, 2t)/(m,\, 4t+1], \qquad (132)$$

$$\text{(C)}\qquad \mu_{2f-1,2g-1} = \sum_{t=f,g}^{t=k} \phi_{2f-1,2t-1}\,\phi_{2g-1,2t-1}\,(4t-2,\, 2t-1)/(m,\, 4t-1], \qquad (133)$$

$$\text{(D)}\qquad \lambda_{2f-1,2g-1} = \sum_{t=f,g}^{t=k} \theta_{2f-1,2t-1}\,\theta_{2g-1,2t-1}\,(4t-2,\, 2t-1)/(m,\, 4t-1], \qquad (134)$$

$$\text{(E)}\qquad \mu_{2f-2,2g-2} = \sum_{t=f,g}^{t=k} \phi_{2f-2,2t-2}\,\phi_{2g-2,2t-2}\,(4t-4,\, 2t-2)/(m,\, 4t-3]. \qquad (135)$$

To find the m.s.e. of the improved value of any l.c. of the differences, or the m.p.e. of two such l.cc., we apply (II.) of § 6, as in § 19 (ii.). Thus, for $m = 2n+1$,

m.p.e. of $b_0v_0+b_1\mu\delta v_0+b_2\delta^2v_0+\ldots+b_{2k}\delta^{2k}v_0$ and $c_0v_0+c_1\mu\delta v_0+c_2\delta^2v_0+\ldots+c_{2k}\delta^{2k}v_0$

$$= \sum_{t=0}^{t=k} \left\{ \sum_{f=0}^{f=t} b_{2f}\,\theta_{2f,2t} \right\} \left\{ \sum_{g=0}^{g=t} c_{2g}\theta_{2g,2t} \right\} (4t,\, 2t)/(m,\, 4t+1]$$

$$+ \sum_{t=1}^{t=k} \left\{ \sum_{f=1}^{f=t} b_{2f-1}\phi_{2f-1,2t-1} \right\} \left\{ \sum_{g=1}^{g=t} c_{2g-1}\phi_{2g-1,2t-1} \right\} (4t-2,\, 2t-1)/(m,\, 4t-1].$$

(viii.) The λ's having been found, the ε's are then given in terms of the σ's by (40). Using the expressions for the σ's given in (26) and (31)–(35), we get the following :—

$$\text{(A)}\qquad \Delta^f v_0 = \sum_{g=0}^{g=j} \lambda_{f,g}\,\Sigma''^{g+1}u_g \qquad (136)$$

$$= \left[ L\left\{ \sum_{g=0}^{g=j} (-)^g\,\lambda_{f,g}\,\Sigma^{g+1}u_g\,;\ \Sigma u_t \right\} \right]_{t=0}^{t=m}; \qquad (136\text{A})$$

$$\text{(B)}\qquad \delta^{2f} v_0 = \left[ L\left\{ \sum_{g=0}^{g=k} \lambda_{2f,2g}\,\mu\sigma^{2g+1}u_0\,;\ \sigma u_t \right\} \right]_{t=-\frac{1}{2}m}^{t=\frac{1}{2}m}; \qquad (137)$$

$$\text{(C)}\qquad \mu\delta^{2f-1} v_0 = \left[ L\left\{ -\sum_{g=1}^{g=k} \mu_{2f-1,2g-1}\,\sigma^{2g}u_0\,;\ \sigma u_t \right\} \right]_{t=-\frac{1}{2}m}^{t=\frac{1}{2}m}; \qquad (138)$$

$$\text{(D)} \qquad \delta^{2f-1}v_{\frac{1}{2}} = \left[L\left\{-\sum_{g=1}^{g=k}\lambda_{2f-1,2g-1}\mu\sigma^{2g}u_{\frac{1}{2}}\,;\ \sigma u_{t+\frac{1}{2}}\right\}\right]_{t=-\frac{1}{2}m}^{t=\frac{1}{2}m}; \quad . \quad . \quad . \quad (139)$$

$$\text{(E)} \qquad \mu\delta^{2f-2}v_{\frac{1}{2}} = \left[L\left\{\sum_{g=1}^{g=k}\mu_{2f-2,2g-2}\sigma^{2g-1}u_{\frac{1}{2}}\,;\ \sigma u_{t+\frac{1}{2}}\right\}\right]_{t=-\frac{1}{2}m}^{t=\frac{1}{2}m}. \quad . \quad . \quad . \quad . \quad (140)$$

The expression in (136) differs slightly from that given for $\Delta^f v_0$ in "Fitting," (15) and § 5; see Appendix III.

(ix.) Finally, we want to express the $\epsilon$'s in terms of the $u$'s. This is done by means of the general theorem in § 10 (iii.), that the coefficients of the $y$'s in any $\epsilon$ are related to the m.pp.e. of this $\epsilon$ and the $\epsilon$'s in the same way that the $v$'s are related to the $\epsilon$'s. Thus we find that—

$$\text{(A)} \qquad \text{If} \quad \Delta^f v_0 = \sum_{r=0}^{r=m} p_r u_r, \quad \text{then} \quad p_r = \sum_{g=0}^{g=j}(r,\,g)\,\lambda_{f,g}\,; \quad . \quad . \quad . \quad . \quad . \quad . \quad (141)$$

$$\text{(B)} \qquad \text{If} \quad \delta^{2f}v_0 = \sum_{r=-n}^{r=n} p_r u_r, \quad \text{then} \quad p_r = \sum_{g=0}^{g=k}[r,\,2g)\,\lambda_{2f,2g}\,; \quad . \quad . \quad . \quad . \quad . \quad (142)$$

$$\text{(C)} \qquad \text{If} \quad \mu\delta^{2f-1}v_0 = \sum_{r=-n}^{r=n} p_r u_r, \quad \text{then} \quad p_r = \sum_{g=1}^{g=k}(r,\,2g-1]\,\mu_{2f-1,2g-1}\,; \quad . \quad . \quad . \quad (143)$$

$$\text{(D)} \qquad \text{If} \quad \delta^{2f-1}v_{\frac{1}{2}} = \sum_{r=-n+1}^{r=n} p_r u_r, \quad \text{then} \quad p_r = \sum_{g=1}^{g=k}[r-\tfrac{1}{2},\,2g-1)\,\lambda_{2f-1,2g-1}\,; \quad . \quad (144)$$

$$\text{(E)} \qquad \text{If} \quad \mu\delta^{2f-2}v_{\frac{1}{2}} = \sum_{r=-n+1}^{r=n} p_r u_r, \quad \text{then} \quad p_r = \sum_{g=1}^{g=k}(r-\tfrac{1}{2},\,2g-2]\,\mu_{2f-2,2g-2}. \quad . \quad . \quad (145)$$

For a comparison of these formulæ with those given in "Fitting," see Appendix IV.

22. *Extent of Improvement (Central Differences).*—A question of practical importance is the extent to which the use of these formulæ actually reduces the m.s.e. of some selected quantity, such as, for the cases marked (B), $u_0$ or $\delta^{2f}u_0$. The m.ss.e. of the various improved values are found from (131)–(135), by putting $g = f$. Comparing these with the m.ss.e. of the original values, for the central-difference formulæ (which are the important ones for practical use), we obtain the following:—

$$\text{(B)} \quad \frac{\text{m.s.e. of } \delta^{2f}v_0}{\text{m.s.e. of } \delta^{2f}u_0} = \frac{1}{(4f,\,2f)}\sum_{t=f}^{t=k}\left\{\frac{\{2t,\,2f\}\,[\frac{1}{2}m,\ 2t+1)}{\{2t,\ 2t\}\,[\frac{1}{2}m,\,2f+1)}\right\}^2\frac{(4t,\ 2t)}{(m,\ 4t+1]}$$

$$= \frac{1}{(m,\ 4f+1]}\sum_{t=f}^{t=k}\frac{4t+1}{4f+1}\left\{\frac{\{2t,\ 2f\}}{\{2f,\,2f\}}\right\}^2$$

$$\frac{\{m^2-(2f+1)^2\}\,\{m^2-(2f+3)^2\}\,\ldots\,\{m^2-(2t-1)^2\}}{\{m^2-(2f+2)^2\}\,\{m^2-(2f+4)^2\}\,\ldots\,\{m^2-(2t)^2\}}; \quad (146)$$

$$(\text{C})\quad \frac{\text{m.s.e. of } \mu\delta^{2f-1}v_0}{\text{m.s.e. of } \mu\delta^{2f-1}u_0} = \frac{4f}{(4f-2,\ 2f-1)} \sum_{t=f}^{t=k} \left\{\frac{\{2t-1,\ 2f-1\}}{\{2t-1,\ 2t-1\}}\,\frac{(\tfrac{1}{2}m,\ 2t]}{(\tfrac{1}{2}m,\ 2f]}\right\}^2 \frac{(4t-2,\ 2t-1)}{(m,\ 4t-1]}$$

$$= \frac{4f}{(m,\ 4f-1]} \sum_{t=f}^{t=k} \frac{4t-1}{4f-1} \left\{\frac{\{2t-1,\ 2f-1\}}{\{2f-1,\ 2f-1\}}\right\}^2$$

$$\frac{\{m^2-(2f+1)^2\}\ \{m^2-(2f+3)^2\}\ \dots\ \{m^2-(2t-1)^2\}}{\{m^2-(2f)^2\}\ \{m^2-(2f+2)^2\}\ \dots\ \{m^2-(2t-2)^2\}}\,; \qquad (147)$$

$$(\text{D})\quad \frac{\text{m.s.e. of } \delta^{2f-1}v_{\frac{1}{2}}}{\text{m.s.e. of } \delta^{2f-1}u_{\frac{1}{2}}} = \frac{1}{(4f-2,\ 2f-1)} \sum_{t=f}^{t=k} \left\{\frac{\{2t-1,\ 2f-1\}}{\{2t-1,\ 2t-1\}}\,\frac{[\tfrac{1}{2}m,\ 2t)}{[\tfrac{1}{2}m,\ 2f)}\right\}^2 \frac{(4t-2,\ 2t-1)}{(m,\ 4t-1]}$$

$$= \frac{1}{(m,\ 4f-1]} \sum_{t=f}^{t=k} \frac{4t-1}{4f-1} \left\{\frac{\{2t-1,\ 2f-1\}}{\{2f-1,\ 2f-1\}}\right\}^2$$

$$\frac{\{m^2-(2f)^2\}\ \{m^2-(2f+2)^2\}\ \dots\ \{m^2-(2t-2)^2\}}{\{m^2-(2f+1)^2\}\ \{m^2-(2f+3)^2\}\ \dots\ \{m^2-(2t-1)^2\}}\,; \qquad (148)$$

$$(\text{E})\quad \frac{\text{m.s.e. of } \mu\delta^{2f-2}v_{\frac{1}{2}}}{\text{m.s.e. of } \mu\delta^{2f-2}u_{\frac{1}{2}}} = \frac{4f-2}{(4f-4,\ 2f-2)} \sum_{t=f}^{t=k} \left\{\frac{\{2t-2,\ 2f-2\}}{\{2t-2,\ 2t-2\}}\,\frac{(\tfrac{1}{2}m,\ 2t-1]}{(\tfrac{1}{2}m,\ 2f-1]}\right\}^2 \frac{(4t-4,\ 2t-2)}{(m,\ 4t-3]}$$

$$= \frac{4f-2}{(m,\ 4f-3]} \sum_{t=f}^{t=k} \frac{4t-3}{4f-3} \left\{\frac{\{2t-2,\ 2f-2\}}{\{2f-2,\ 2f-2\}}\right\}^2$$

$$\frac{\{m^2-(2f)^2\}\ \{m^2-(2f+2)^2\}\ \dots\ \{m^2-(2t-2)^2\}}{\{m^2-(2f-1)^2\}\ \{m^2-(2f+1)^2\}\ \dots\ \{m^2-(2t-3)^2\}}\,. \qquad (149)$$

23. *Smoothing.*—When we have a table containing a very large number of $u$'s, a common method of procedure is to use a formula involving a limited number of terms and to apply it to successive sets of the $u$'s for the purpose of obtaining a table to be substituted for the original table. Thus we might use a formula involving $2n+1$ terms, and apply it to $u_0, u_1, u_2, \dots u_{2n}$ for finding a new value $w_n$, then to $u_1, u_2, u_3, \dots u_{2n+1}$ for finding a new value $w_{n+1}$, and so on. These values having been obtained, a differenced table would be formed; but, as by hypothesis the true differences of order exceeding $j$ are negligible, the table would only go up to differences of order $j$. There are two cases to be considered.

(i.) If our object is to obtain as accurate values as possible for the $w$'s, consistently with our using only the specified number of $u$'s for each, the most accurate values would be the $v$'s given by the formulæ considered in this and the preceding papers. It should, however, be observed that the differences of the $w$'s are not the same as the $\Delta^f v_0$, $\delta^{2f}v_0$, etc., occurring in those formulæ. Suppose, for instance, that we replace $u_0$ by its improved value $v_0$ obtained by means of the (B) formula involving $u_{-n}$, $u_{-n+1}$, $\dots u_n$, and replace $u_1$ by the improved value $v_1$ obtained in a similar way. The resulting value of $v_1 - v_0$ will involve the $2n+2$ $u$'s from $u_{-n}$ to $u_{n+1}$; but it will not be the same thing as the improved value $v_{\frac{1}{2}}$ obtained by the (D) formula involving these $u$'s, and its m.s.e. will therefore be greater than that of the latter.

(ii.) If our object is to obtain a smooth table of the $w$'s as a whole, we could do this by obtaining as accurate values as possible for the differences of the $w$'s of order $j$.

The formula which would have to be applied to the $u$'s in order to obtain this result can be constructed without difficulty. The important thing to notice is that, if we alter the differences of the $u$'s and then obtain the $w$'s from the altered differences by summation, the resulting values must be such as can be legitimately substituted for the $u$'s. Suppose, for instance, that $j = 2k+1$, and that we use $2n+1$ $u$'s for each $w$. The formula for $w$ will have to be of the form

$$w_0 = u_0 + c_{2k+2}\delta^{2k+2}u_0 + c_{2k+4}\delta^{2k+4}u_0 + \ldots + c_{2n}\delta^{2n}u_0\,;$$

and this will give

$$\delta^{2k+1}w_{\frac{1}{2}} = \delta^{2k+1}u_{\frac{1}{2}} + c_{2k+2}\delta^{4k+3}u_{\frac{1}{2}} + c_{2k+4}\delta^{4k+5}u_{\frac{1}{2}} + \ldots + c_{2n}\delta^{2n+2k+1}u_{\frac{1}{2}}.$$

The problem of determining the $c$'s so that the m.s.e. of $\delta^{2k+1}w_{\frac{1}{2}}$ shall be a minimum is the same as that of determining the coefficients in the improved value of $\delta^{2k+1}u_{\frac{1}{2}}$ for $j = 4k+1$ or $4k+2$, $m$ being $2n+2k+2$; and the solution of this problem is given in § 21. Thus, in terms of sums, (139) gives

$$\delta^{2k+1}w_{\frac{1}{2}} = \left[L\left\{-\sum_{g=1}^{g=2k+1}\lambda_{2k+1,\,2g-1}\mu\sigma^{2g}u_{\frac{1}{2}}\,;\ \sigma u_{t+\frac{1}{2}}\right\}\right]_{t=-(n+k+1)}^{t=n+k+1}.$$

The $\lambda$'s having been found, we shall then have, by summation,

$$w_0 = \left[L\left\{-\sum_{g=1}^{g=2k+1}\lambda_{2k+1,\,2g-1}\mu\sigma^{2g+2k+1}u_0\,;\ \sigma^{2k+2}u_t\right\}\right]_{t=-(n+k+1)}^{t=n+k+1}.$$

The ratio of the m.s.e. of $\delta^{2k+1}w_{\frac{1}{2}}$ to that of $\delta^{2k+1}u_{\frac{1}{2}}$ is given by (148).

## APPENDIX I.—THE CORRELATION-DETERMINANT.

1. The m.p.e. of $A$ and $B$ being denoted by $(A\,;\,B)$, let

$$\Theta \equiv \begin{vmatrix} (A\,;\,A) & (A\,;\,B) & (A\,;\,C) & \ldots \\ (B\,;\,A) & (B\,;\,B) & (B\,;\,C) & \ldots \\ (C\,;\,A) & (C\,;\,B) & (C\,;\,C) & \ldots \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \\ \cdot & \cdot & \cdot & \end{vmatrix}.$$

We call this the *correlation-determinant* of $A$, $B$, $C$, ....

2. The elements of this determinant may be regarded as obtained as follows. We first take a representative collection of $N_A$ values of the error of $A$; $N_A$ being usually indefinitely great. Then, for each of these values, we take a representative collection of $N_B$ values of the error of $B$; the resulting $N_A$ collections will all be alike if the errors of $A$ and of $B$ are independent, but not if they are correlated. This gives $N_AN_B$ combinations of an error of $A$ and an error of $B$. For each of these we take

a representative collection of $N_C$ values of the error of $C$; and so on. Thus finally we shall have $N \equiv N_A N_B N_C \ldots$ combinations of an error of $A$, an error of $B$, &c. Numbering these 1, 2, ... $N$, and denoting the errors of $A$, of $B$, of $C$, ... by $a$, $b$, $c$, ..., the combinations will be $a_1, b_1, c_1, \ldots, a_2, b_2, c_2, \ldots, \ldots a_N, b_N, c_N, \ldots$; and we shall have

$$(A\ ;\ A) = (a_1^2 + a_2^2 + a_3^2 + \ldots + a_N^2)/N,$$
$$(A\ ;\ B) = (a_1b_1 + a_2b_2 + a_3b_3 + \ldots + a_Nb_N)/N,$$
$$\&c.$$

3. Substituting these values in $\Theta$, we find that, if there are $m$ of the quantities $A, B, C, \ldots$,

$$N^m\Theta = \begin{vmatrix} a_1^2 + a_2^2 + a_3^2 + \ldots & a_1b_1 + a_2b_2 + a_3b_3 + \ldots & a_1c_1 + a_2c_2 + a_3c_3 + \ldots & \&c. \\ a_1b_1 + a_2b_2 + a_3b_3 + \ldots & b_1^2 + b_2^2 + b_3^2 + \ldots & b_1c_1 + b_2c_2 + b_3c_3 + \ldots & \&c. \\ a_1c_1 + a_2c_2 + a_3c_3 + \ldots & b_1c_1 + b_2c_2 + b_3c_3 + \ldots & c_1^2 + c_2^2 + c_3^2 + \ldots & \&c. \\ \vdots & \vdots & \vdots & \vdots \end{vmatrix}$$

$$= (a_1b_2c_3\ldots)^2 + (a_1b_2c_4\ldots)^2 + (a_1b_3c_1\ldots)^2 + (a_2b_3c_4\ldots)^2 + \ldots,$$

where $(a_1b_2c_3\ldots)$ denotes $\begin{vmatrix} a_1a_2a_3\ldots \\ b_1b_2b_3\ldots \\ c_1c_2c_3\ldots \\ \vdots \end{vmatrix}$. Hence $\Theta$ is not $= 0$ unless each of the determinants $(abc\ldots)$ is $= 0$. This would be the case, for instance, if $A$ were a constant, so that every $a$ would be 0, or if there were a linear relation connecting the errors of $A, B, C, \ldots$.

4. Let $\Phi$ be the correlation-determinant of $A, B, \ldots P, Q, \ldots$, and $\Psi$ that of $A, B, \ldots P$.

($a$) Suppose that $\Psi = 0$. Then, by § 3 of this Appendix, each of the determinants $(ab \ldots p)$, where $a, b, \ldots p$ are the errors of $A, B, \ldots P$, is 0. But these are the minors of the $q$'s in the determinants $(ab \ldots pq)$; and therefore these latter determinants are 0. Proceeding in this way, we see that the determinants $(ab \ldots pq \ldots)$ are all 0; and therefore $\Phi = 0$.

($b$) Hence, if $\Phi$ is not $= 0$, $\Psi$ is not $= 0$.

## APPENDIX II.—FREQUENCY OF CORRELATED ERRORS.

1. Let $u_0, u_1, u_2, \ldots u_l$ and $y_0, y_1, y_2, \ldots y_l$ be two conjugate sets. Denote the errors of the $u$'s by $\theta_0, \theta_1, \theta_2, \ldots \theta_l$; and let the resulting errors of the $y$'s be $\phi_0, \phi_1, \phi_2, \ldots \phi_l$.

Then, on the assumption of normal correlation of errors, the frequency of joint occurrence of these $\theta$'s is proportional to

$$\exp - \tfrac{1}{2}P,$$

where $P$ is a homogeneous quadratic function of the $\theta$'s. We want to prove that

(i.) the $\phi$'s are the partial differential coefficients of $\frac{1}{2}P$ with regard to the $\theta$'s, and conversely;

(ii.) $P = \theta_0\phi_0 + \theta_1\phi_1 + \theta_2\phi_2 + \ldots + \theta_l\phi_l$;

(iii.) $P = \psi_{0,0}\theta_0^2 + 2\psi_{0,1}\theta_0\theta_1 + \psi_{1,1}\theta_1^2 + \ldots + \psi_{l,l}\theta_l^2$, where $\psi_{f,g}$ is the m.p.e. of $y_f$ and $y_g$; and similarly

(iv.) $P = \pi_{0,0}\phi_0^2 + 2\pi_{0,1}\phi_0\phi_1 + \pi_{1,1}\phi_1^2 + \ldots + \pi_{l,l}\phi_l^2$, where $\pi_{f,g}$ is the m.p.e of $u_f$ and $u_g$.

2. Suppose that

$$P = a_{0,0}\theta_0^2 + 2a_{0,1}\theta_0\theta_1 + a_{1,1}\theta_1^2 + \ldots + a_{l,l}\theta_l^2;$$

and let us, without making any assumption of conjugacy, write ($f = 0, 1, 2, \ldots l$)

$$\begin{aligned} y_f &\equiv a_{f,0}u_0 + a_{f,1}u_1 + a_{f,2}u_2 + \ldots + a_{f,l}u_l, \\ \phi_f &\equiv \text{error of } y_f \\ &= a_{f,0}\theta_0 + a_{f,1}\theta_1 + a_{f,2}\theta_2 + \ldots + a_{f,l}\theta_l \\ &= \tfrac{1}{2}dP/d\theta_f. \end{aligned}$$

Then, writing the subscripts in the order $f, 0, 1, 2, \ldots l$,

$$\begin{aligned} P &= a_{f,f}\theta_f^2 + 2a_{f,0}\theta_f\theta_0 + a_{0,0}\theta_0^2 + \ldots + a_{l,l}\theta_l^2 \\ &= \phi_f^2/a_{f,f} + Q, \end{aligned}$$

where $Q$ does not contain $\theta_f$.

3. The mean value of $\phi_f\theta_g$ is $N_g/D$, where

$$N_g \equiv \iiint \ldots \int \phi_f\theta_g \exp - \tfrac{1}{2}P \,.\, d\theta_f\, d\theta_0\, d\theta_1 \ldots d\theta_l,$$

$$D \equiv \iiint \ldots \int \exp - \tfrac{1}{2}P \,.\, d\theta_f\, d\theta_0\, d\theta_1 \ldots d\theta_l,$$

the integration being in each case from $-\infty$ to $\infty$. If we write

$$\psi \equiv \phi_f/\sqrt{a_{f,f}},$$

then

$$N_g = \iiint \ldots \int \psi\theta_g \exp - \tfrac{1}{2}\psi^2 \,.\, \exp - \tfrac{1}{2}Q \,.\, d\psi\, d\theta_0\, d\theta_1 \ldots d\theta_l.$$

($a$). First, suppose that $g$ is not $= f$. Then, integrating with regard to $\psi$,

$$N_g = 0,$$

(*b*) Next, suppose that $g = f$. Then

$$N_f = 1/a_{f,f} \cdot \iiint \dots \int \phi_f (\phi_f - a_{f,0}\theta_0 - a_{f,1}\theta_1 - \dots - a_{f,l}\theta_l) \exp -\tfrac{1}{2}P \,.\, d\theta_f \, d\theta_0 \, d\theta_1 \dots d\theta_l$$

$$= 1/a_{f,f} \cdot \iiint \dots \int \phi_f^2 \exp -\tfrac{1}{2}P \,.\, d\theta_f \, d\theta_0 \, d\theta_1 \dots d\theta_l,$$

by (*a*). Hence

$$N_f = 1/\sqrt{a_{f,f}} \cdot \iiint \dots \int \psi^2 \exp -\tfrac{1}{2}\psi^2 \,.\, \exp -\tfrac{1}{2}Q \,.\, d\psi \, d\theta_0 \, d\theta_1 \dots d\theta_l$$

$$= \sqrt{(2\pi/a_{f,f})} \cdot \iint \dots \int \exp -\tfrac{1}{2}Q \,.\, d\theta_0 \, d\theta_1 \dots d\theta_l.$$

Also

$$D = 1/\sqrt{a_{f,f}} \cdot \iiint \dots \int \exp -\tfrac{1}{2}\psi^2 \,.\, \exp -\tfrac{1}{2}Q \,.\, d\psi \, d\theta_0 \, d\theta_1 \dots d\theta_l$$

$$= \sqrt{(2\pi/a_{f,f})} \cdot \iint \exp -\tfrac{1}{2}Q \,.\, d\theta_0 \, d\theta_1 \dots d\theta_l.$$

Hence

$$N_f/D = 1.$$

4. Hence the $y$'s are related to the $u$'s in such a way that

$$\text{m.p.e. of } y_f \text{ and } u_g = 0 \; (g \neq f) \text{ or } 1 \; (g = f)\,;$$

and therefore the $u$'s and the $y$'s are conjugate sets; which proves (i.). It follows that

$$a_{f,g} = \text{m.p.e. of } y_f \text{ and } y_g.$$

This proves (iii.); and (iv.) is the similar result which we should have obtained by expressing $P$ in terms of the $\phi$'s. Also

$$\begin{aligned} P &= a_{0,0}\theta_0^2 + 2a_{0,1}\theta_0\theta_1 + a_{1,1}\theta_1^2 + \dots + a_{l,l}\theta_l^2 \\ &= \theta_0 (a_{0,0}\theta_0 + a_{0,1}\theta_1 + a_{0,2}\theta_2 + \dots + a_{0,l}\theta_l) \\ &\quad + \theta_1 (a_{1,0}\theta_0 + a_{1,1}\theta_1 + a_{1,2}\theta_2 + \dots + a_{1,l}\theta_l) \\ &\quad + \dots \\ &\quad + \theta_l (a_{l,0}\theta_0 + a_{l,1}\theta_1 + a_{l,2}\theta_2 + \dots + a_{l,l}\theta_l) \\ &= \theta_0\phi_0 + \theta_1\phi_1 + \theta_2\phi_2 + \dots + \theta_l\phi_l\,; \end{aligned}$$

which proves (ii.).

## APPENDIX III.—IMPROVED ADVANCING DIFFERENCES IN TERMS OF SUMS.

The expression for $\Delta^f v_0$ given by (136) differs from that given in (15) and § 5 of "Fitting," in that it involves $\Sigma'' u_0$, $\Sigma''^2 u_1$, $\Sigma''^3 u_2, \dots$, instead of $\Sigma'' u_0$, $\Sigma''^2 u_0$, $\Sigma''^3 u_0, \dots$.

The new expressions are more convenient for calculation and for tabulation, since the coefficients are rather smaller and are symmetrically placed about a diagonal. For $j = 2$, $m = 13$, for instance, the formulæ given by "Fitting," § 5 (ii.), are

$$1001v_0 = +693S_1-198S_2+22S_3,$$

$$1001\Delta v_0 = -231S_1+88S_2-11S_3,$$

$$1001\Delta^2 v_0 = +35S_1-15S_2+2S_3,$$

where $S_1 \equiv \Sigma'' u_0$, $S_2 \equiv \Sigma''^2 u_0$, $S_3 \equiv \Sigma''^3 u_0$. If we write $\Sigma_1 \equiv \Sigma'' u_0$, $\Sigma_2 \equiv \Sigma''^2 u_1$, $\Sigma_3 \equiv \Sigma''^3 u_2$, these become (by (136), or by writing $S_1 = \Sigma_1$, $S_2 = \Sigma_1 + \Sigma_2$, $S_3 = \Sigma_1+2\Sigma_2+\Sigma_3$)

$$1001v_0 = +517\Sigma_1-154\Sigma_2+22\Sigma_3,$$

$$1001\Delta v_0 = -154\Sigma_1+66\Sigma_2-11\Sigma_3.$$

$$1001\Delta^2 v_0 = +22\Sigma_1-11\Sigma_2+2\Sigma_3.$$

The symmetry of the coefficients is due to the fact that

$$\text{co. } \Sigma''^{g+1}u_g \text{ in } \Delta^f v_0 = \lambda_{f,g} = \lambda_{g,f} = \text{co. } \Sigma''^{f+1}u_f \text{ in } \Delta^g v_0.$$

For any particular value of $m$ there will be only $\frac{1}{2}(j+1)(j+2)$ coefficients to be tabulated, instead of $(j+1)^2$.

## APPENDIX IV.—FORMULÆ IN TERMS OF $u$'S.

(i.) Formulæ for $\Delta^f v_0$, &c., in terms of the $u$'s have already been given in (15$a$), (21), (22), (29), and (28) of "Fitting"; and the results in (141)–(145) of the present paper can be checked by comparing the different expressions for the coefficients of the $u$'s. We should require to use the following identities:—

$$(r+h,\ h) = (r,\ 0)\ (h,\ 0)+(r,\ 1)\ (h,\ 1)+(r,\ 2)\ (h,\ 2)+\ \ldots,$$

$$(r,\ 2h] = (0,\ 2h]+[r,\ 2)\ (0,\ 2h-2]+[r,\ 4)\ (0,\ 2h-4]+\ \ldots,$$

$$[r,\ 2h-1) = (r,\ 1]\ [\pm\tfrac{1}{2},\ 2h-2)+(r,\ 3]\ [\pm\tfrac{1}{2},\ 2h-4)+(r,\ 5]\ [\pm\tfrac{1}{2},\ 2h-6)+\ \ldots,$$

$$(r-\tfrac{1}{2},\ 2h-1] = [r-\tfrac{1}{2},\ 1)\ (0,\ 2h-2]+[r-\tfrac{1}{2},\ 3)\ (0,\ 2h-4]+[r-\tfrac{1}{2},\ 5)\ (0,\ 2h-6]+\ \ldots,$$

$$[r-\tfrac{1}{2},\ 2h-2) = [\pm\tfrac{1}{2},\ 2h-2)+(r-\tfrac{1}{2},\ 2]\ [\pm\tfrac{1}{2},\ 2h-4)+(r-\tfrac{1}{2},\ 4]\ [\pm\tfrac{1}{2},\ 2h-6)+\ \ldots.$$

(ii.) Taking, for instance, the formula for $\delta^{2f}v_0$ when $m = 2n+1$, (21) of "Fitting" gives (replacing $t$ by $f$)

$$(p_r)_{2k} = (-)^f \tfrac{1}{2}m\,\frac{\{2k+1,\ 2f+1\}}{[\frac{1}{2}m,\ 2f+1)}\sum_{h=0}^{h=k}(-)^h\frac{(f+1)(h+1)}{f+h+\frac{1}{2}}\,\frac{\{2k+1,\ 2h+1\}\,(r.\ 2h]}{(\frac{1}{2}m,\ 2h+1]}.$$

Hence we find that

$$(p_r)_{2k}-(p_r)_{2k-2} = (-)^f\tfrac{1}{2}m\,(2k+\tfrac{1}{2})\frac{\{2k,\ 2f\}}{[\frac{1}{2}m,\ 2f+1)}\sum_{h=0}^{h=k}(-)^h\frac{\{2k,\ 2h\}\ (r,\ 2h]}{(\frac{1}{2}m,\ 2h+1]},$$

$$(p_r)_{2k} = (-)^f\tfrac{1}{2}m\sum_{t=f}^{t=k}(2t+\tfrac{1}{2})\frac{\{2t,\ 2f\}}{[\frac{1}{2}m,\ 2f+1)}\sum_{h=0}^{h=t}(-)^h\frac{\{2t,\ 2h\}\ (r,\ 2h]}{(\frac{1}{2}m,\ 2h+1]};$$

and therefore

$$(\delta^{2g}p_0)_{2k} = (-)^{f+g}\tfrac{1}{2}m\sum_{t=f}^{t=k}(2t+\tfrac{1}{2})\frac{\{2t,\ 2f\}}{[\frac{1}{2}m,\ 2f+1)}\frac{\{2t,\ 2g\}}{(\frac{1}{2}m,\ 2g+1]}\Theta,$$

where

$$\Theta = F\{-t+g,\ g+t+\tfrac{1}{2},\ \tfrac{1}{2}\,;\ 1,\ -\tfrac{1}{2}m+g+1,\ \tfrac{1}{2}m+g+1\}$$

$$= \frac{(\frac{1}{2}m,\ 2g+1]\ [\frac{1}{2}m,\ 2t+1)}{(\frac{1}{2}m,\ 2t+1]\ [\frac{1}{2}m,\ 2g+1)}.$$

This expression for $(\delta^{2g}p_0)_{2k}$ will be found to be equal to $\lambda_{2f,2g}$ as given by (132) of the present paper, so that the formula in "Fitting" agrees with (142).

# VIII. *Tidal Friction in Shallow Seas.*

*By* HAROLD JEFFREYS, *M.A., D.Sc., Fellow of St. John's College, Cambridge.*

*Communicated by Sir* NAPIER SHAW, *F.R.S.*

Received April 7,—Read June 24, 1920.

THE astronomical importance of the dissipation of energy that goes on in shallow seas has been shown by G. I. TAYLOR'S recent estimate* of the amount in the Irish Sea, which is enough to account for about one-fiftieth of the secular acceleration of the moon. It also produces a considerable effect on the tides themselves, and there are probably many places where it must be taken into account before any satisfactory theory of the local tides, or even their empirical prediction, can be achieved. It is indeed very well known that there are bays and straits where the height of the tides, or the speed of the currents, or both, are greater than in the Irish Sea, and a careful examination of such places, with a view to finding the dissipation in them, is needed. There are other places where the dissipation for an equal area is less than in the Irish Sea, but which may actually contribute much more altogether on account of their greater size. The object of this paper is to discuss what regions are capable of producing notable parts of the secular acceleration; to estimate as accurately as possible from the data available the dissipation in these; and to compare this with that calculated from the secular acceleration, so as to find out whether it is necessary to assume the existence of any other important cause to account for the latter.

The horizontal force of the skin friction of water over the sea bottom is $0{\cdot}002\rho V^2$ dynes per square centimetre, where $\rho$ is measured in grammes per cubic centimetre and V in centimetres per second. The difficulty of the problem is in the estimation of V. The available observations of the velocities of tidal currents are given in the Admiralty Sailing Directions; but they are never uniformly distributed, and are usually confined to the neighbourhood of the coasts, and they must be supplemented by theory before the velocities remote from the coast can be found. A few theoretical considerations that have been found useful in this process will now be mentioned.

Take first the case of a bay or strait long in comparison with its width, and consider a wave entering it whose period is much longer than the time needed for a

---

* 'Phil. Trans.,' A, vol. 220, pp. 1–33, 1919.

[Published November 5, 1920

wave of tidal type to cross. This time is known to be independent of the period of the disturbance, depending only on the width and depth. Such a wave is reflected from one side to the other again and again before it reaches the other end. It is also known that the transverse velocity at the sides is zero, since water cannot cross a rigid boundary. Thus if we compare two points on opposite sides of the channel, we know that the times of arrival of the wave at them differ by a small fraction of a period; and since the transverse velocity at both is zero, it cannot be great at any intermediate point, for that would contradict the hypothesis that the wave-length is much greater than the width of the channel. The transverse velocity may accordingly be neglected in problems of this class.

If the period of the entering wave is of the same order of magnitude as the time needed to cross the channel, we can no longer infer that the transverse velocity is much less than the longitudinal one. This case seldom or never arises. The velocity of a tidal wave is $(g\mathrm{D})^{\frac{1}{2}}$ where $g$ is the intensity of gravity and D is the depth; and if the water was only 20 fathoms deep a tidal wave would in 12 hours travel 700 km., which is far greater than the width of almost any channel whose length is much greater than its width. Where the width is greater, the depth also is always greater, so that the above argument always holds in long channels of whatever size.

If now $x$ be the distance of a point from the entrance to the channel, $y$ the distance from the side, $u$ the longitudinal velocity of a particle there, and $\eta$ the height of the free surface above its undisturbed position, the equations of motion of the particle are

$$\frac{du}{dt} = -\frac{g\partial\eta}{\partial x} - \text{a term due to friction}$$

$$2\omega u = -g\frac{\partial\eta}{\partial y}$$

where $\omega$ is the component of the earth's angular velocity of rotation about the vertical at the point. The equation of continuity is

$$\frac{\partial}{\partial x}(\mathrm{D}u) = -\frac{\partial\eta}{\partial t}.$$

From this and the second equation of motion we deduce at once

$$\frac{\partial\eta}{\partial x} = \frac{\partial\eta_0}{\partial x} + \frac{2\omega}{g}\int\frac{1}{\mathrm{D}}\left(\frac{\partial\eta}{\partial t} + u\frac{\partial\mathrm{D}}{\partial x}\right)dy$$

where $\eta_0$ is the value of $\eta$ at the side. In this we see from the conditions that the channel is narrow and the depth slowly varying along it that the first term is much greater than the others. Accordingly $\frac{\partial\eta}{\partial x}$ is the same for all particles in the same cross-section of the channel, and the first equation of motion then shows that the same is true of $u$ if the friction is small. Thus in such cases the velocity is the same at all points in the same cross-section, so that observations made at the side will be correct for points in the middle. If friction is great this result must be modified, since for the same velocity the frictional force is independent of the depth of the water, whereas the mass affected is proportional to the depth. Thus friction has more influence in reducing the velocities in shallow water than in deep water. Hence when the channel is shallow at the sides and deep in the middle the velocity will be least at the sides. The ratio of the frictional term to the first term is $0{\cdot}002\ u/2\mathrm{D}'\Omega$, where $\mathrm{D}'$ is the depth at the shallow part. When this is 10 fathoms and the velocity 1 knot the fraction is 0·25, so that this effect is then appreciable ; and when the depths are less and the velocities greater, the influence of friction may increase to such an extent as to dominate the whole character of the motion. When this occurs, the velocity will always be in the direction of decreasing pressure, and inertia may be neglected.

The last result may appear to contradict the general principle that "still waters run deep." There is no real contradiction, however, for the problems referred to are different. The above argument deals with the differences between the velocities at places in the same cross-section of the channel, whereas the proverb concerns rivers whose depth varies *along* them, and in such cases the motion is naturally slowest where the depth is greatest, since the amount of water crossing any section in a given time must be the same. It also has an important and well-known application in bays of varying depth and width, such as the Bay of Fundy. If, for instance, the bay is very long, and these quantities change only by small fractions of themselves per wave-length, it can be shown* that the height of the wave at any point is proportional to $b^{-\frac{1}{2}}\mathrm{D}^{-\frac{1}{4}}$, and the velocity of the water to $b^{-\frac{1}{2}}\mathrm{D}^{-\frac{3}{4}}$, where $b$ is the width at the surface. The rate of dissipation of energy across any section is proportional to $bu^3$ or to $b^{-\frac{1}{2}}\mathrm{D}^{-\frac{9}{4}}$. It therefore increases slowly as the channel becomes narrower and much more quickly as it becomes shallower. When the depth and width vary much within a wave-length these results cease to be useful approximations, but the tendency for the height of the tide and the velocity of the tidal current to increase as the channel becomes narrower and shallower remains. Thus in such places we often find very high tides and strong tidal currents. Apparently, however, their limited area prevents the dissipation in them from being as great as that in larger places with less violent currents (at least, if the Bay of Fundy may be regarded as typical of them).

The widths of most actual bays are, however, comparable with their lengths, and in these it is generally a matter of some difficulty to settle whether we can treat the recorded currents as a fair sample of the whole. The amplitude of the tide in mid-ocean is only about a foot, but in the shallow water around the coasts it is magnified to several feet, and the tidal currents are increased correspondingly. Where the shore is fairly open and regular in outline, like most of the coast of Africa, it is not possible to find the dissipation along it, for there are no data to show how far out the currents extend. In partly enclosed regions, however, it is frequently possible to interpolate between the records made on opposite sides. A serious difficulty may arise if the depth of the sea is very different at different points within it, for this may destroy the possibility of interpolation, and therefore we must always examine the soundings for any great variation. Ordinarily we should not expect much variation in the velocities, for such places are intermediate in character between narrow channels and the open shore, and therefore the currents in them may be expected to show some increase in shallow water, but not so much as would be caused by a proportional decrease in depth along a narrow bay or in approaching the open shore. In shallow water also friction may, and often does, neutralize the magnification that would occur in its absence.

---

* LAMB, 'Hydrodynamics,' p. 258.

One other fact may be noted. In shallow bays the difficulties of navigation may be great, and navigators avoid them if possible by choosing a harbour near the entrance. Thus observations of currents are most numerous about the entrances, and often at the very places where the currents, and consequently the dissipation, are greatest there are insufficient observations to give a satisfactory estimate.

Great care must be taken in dealing with observations among islands, straits, and shoals. When the passage of a tidal current is obstructed by a shallow of small horizontal extent, part of it goes round the shoal and part over the top. The influence of this on the main current is of course small, but on the top of the shoal and in its immediate neighbourhood the velocities may be much increased, for much the same reason as accounts for the greater speed of a river where it is shallow. On the other hand the increase in the influence of friction may greatly reduce the currents, and shoals often afford in this way an important shelter from tidal currents to the deeper water behind them. This is particularly noticeable at some points on the Korean side of the Yellow Sea. Thus observations of currents taken at lightships and buoys over shoals whose dimensions are all much smaller than those of the main bay or channel must be regarded as giving no reliable estimate of the main current. Small islands also require examination before the records obtained are accepted. If one is surrounded by a shoal they are of course untrustworthy; but if deep soundings are found within a few miles of it, they will probably give a very good idea of the main current, which will, especially in a wide channel, be fully as useful in our investigation as the results of observations at the sides. Straits are in a different position. When the tides in two seas or even oceans are in widely different phases, a large head may be produced between the two ends of a strait connecting them, so that a swift tidal current will flow along the strait. In no circumstances, however, can this give any indication of the currents in the seas, for it is produced by the tide heights, and not directly by the currents. Such currents may attain very great velocities, as in the

Magellan Strait and Smith Sound, and when the area of the strait is not insignificant the dissipation may be an important part of that in the seas as a whole.

## EUROPEAN SEAS.

### 1. *The Irish Sea.*

This sea has been discussed in detail by TAYLOR. The rate of dissipation is found to be 1040 ergs per square centimetre per second, or $4 \cdot 1 \times 10^{17}$ ergs per second in all, on an average at spring tides. This result is based on the law that the rate is proportional to the cube of the velocity. The Irish Sea is remarkable in that the maximum current occurs nearly at high water, whereas in ordinary places the water is nearly slack then; though other examples will be given later in this paper. This affords the most favourable conditions for an accurate estimate of the rate at which energy enters the sea; to this TAYLOR added the rate at which the moon's attraction does work on the sea, and from the fact that all this energy must be dissipated in the course of a period (for if it were not, there would be a continual increase of energy in the Irish Sea) he found that the mean rate of dissipation at spring tides was $6 \cdot 0 \times 10^{17}$ ergs per second. This estimate is probably more accurate than the other, as the data involved are obtained from observations in St. George's Channel, supplemented by an accurate theory; but the former is based on an average of the velocities in the Irish Sea itself, which are more difficult to determine.

### 2. *The English Channel.*

On an average the tidal currents in the English Channel at springs reach about 2·5 knots. The speed is greatest towards the Straits of Dover and least at the entrance to the Channel, and enough data are available in the Admiralty publication, 'The Tides and Tidal Streams of the British Isles' to give a very accurate estimate of the total dissipation if this were required in the present problem; but as the errors introduced by using only a rough approximation in this case are far less than those involved in the best data referring to regions with far larger dissipations, accuracy is not here required.

The rate of dissipation is $0 \cdot 002 \rho V^3$ ergs per square centimetre per second. Here $\rho$ is practically 1; and one knot is 51·5 cm. per second. Thus the dissipation per square centimetre for a velocity of one knot is 274 ergs per second, and that per square kilometre is $2 \cdot 74 \times 10^{12}$ ergs per second. The area of the English Channel is about 60,000 sq. km., so that the dissipation when the currents are flowing fastest is $2 \cdot 74 \times 10^{12} \times 6 \times 10^4 \times 2 \cdot 5^3$, or $2 \cdot 5 \times 10^{18}$ ergs per second. This is of course a maximum, while the value obtained for the Irish Sea is the mean over a period; the average rates of dissipation in the two places are perhaps not very different.

### 3. *The North Sea.*

A satisfactory estimate of the dissipation in the North Sea is practically impossible. Velocities up to over 3 knots are recorded here and there, but all the observations are in the coastal region, which is very much complicated by shoals. The maximum in the outermost part of the Moray Firth is about 1·1 knot, and this is probably fairly typical of the whole of the North Sea. Taking the area to be $5 \times 10^5$ sq. km. and adopting the above value of the velocity, we see that the maximum dissipation is of the order of $1{\cdot}8 \times 10^{18}$ ergs per second.

### 4. *Other European Waters.*

In the Mediterranean there is probably little or no dissipation of tidal energy, for the Atlantic tidal wave can only enter through the very narrow Straits of Gibraltar, and partly for this reason and partly on account of the great length and considerable depth of the sea there is very little tidal movement in it. The same argument applies to the Baltic, for the entrance through the Kattegat is largely blocked up by the Danish islands, so that little water can enter to produce a tide. The Bay of Biscay is mostly too deep to have any important current, while the White Sea is too small and landlocked to give as much dissipation as the Irish Sea.

The average dissipation in a period is $4/3\pi$ of the maximum. If we find the maximum for the Irish Sea on this basis, we obtain for the total dissipation in European waters when the spring tide currents are flowing strongest about $6{\cdot}0 \times 10^{18}$ ergs per second; the average at spring tides is $2{\cdot}4 \times 10^{18}$ ergs per second.

## ASIATIC SEAS.

It has already been pointed out that the tidal currents in mid-ocean are insufficient to give any important dissipation.* Accordingly we need consider only those places where the currents are very much magnified by great decreases in depth. On referring to a physical map of Asia it is at once seen that the places around the coast where the depth is less than 100 fathoms are the Straits of Malacca, the South China Sea (with the Java Sea), the Gulfs of Siam and Tongking, the Yellow Sea, the Persian Gulf, and parts of the Seas of Japan and Okhotsk and the Bering Sea. These regions will be dealt with separately. The Persian Gulf may be omitted at once, as its narrow entrance prevents the tide from being great.

### 1. *The South China Sea.*

This sea is in the form of a letter T. The middle stroke points north-east and lies between Annam and Southern China on the one side, and Borneo and the Philippines on the other. The two side-pieces are the Gulf of Siam and the Java Sea. The data for it are obtained from the 'China Sea Pilot,' volumes 3 and 4, except the depths, which are taken from the Admiralty Charts. The tides are affected by a large diurnal inequality due to the inclination of the Equator to the ecliptic. When the moon is north of the Equator it tends to raise two tidal protuberances in the ocean, one exactly below it and the other exactly opposite to it. Owing to the earth's rotation each of these moves round the earth once a day, keeping the same distance north or south of the Equator. Thus if a place is not on the Equator, they pass at different distances from it, so that the two tides in the lunar day are unequal in height. The variation in the level of the water thus caused can be described as a semi-diurnal change, on which a diurnal change is superposed.

---

* 'Philosophical Magazine,' May, 1920, vol. 39, pp. 578–586.

Now the rate of travel of a tidal wave is practically independent of its period, but if the depth and the form of the coast are such that the waves starting from the north and south sides of the Equator take different times to reach the place of observation, their combined effect may be remarkable. In particular, if the wave from the south arrives a quarter of a lunar day after or before the other, the semi-diurnal part of the one wave will correspond to high water while that of the other corresponds to low water, and if the amplitudes are equal the two will neutralize each other. In other words, there will be a node of the semi-diurnal tide. The diurnal parts, however, will not neutralize each other, their phases being only a quarter of a period apart. Thus at such a place there will be a diurnal tide and no semi-diurnal tide. Several places are known where there is only one high water in each lunar day; among them are parts of the South China Sea, the Gulf of Carpentaria, and Bering Sea. The tides in these require special discussion before they can be considered in the present problem, because the diurnal tide depends essentially on the inclination of the Equator to the plane of the moon's orbit and would not exist if this were zero. The dissipation of energy in it must therefore arise from the motion of the moon in declination and not in right ascension, and will affect mainly the inclinations of the Equator and the moon's orbit to the ecliptic, while producing little effect on the earth's rotation and the motion of the moon in longitude. In discussing the secular acceleration of the moon it can therefore be ignored. If observations of the diurnal tide in the places where there are two high waters in the day were more numerous it might be possible to determine the dissipation in it, and from it the secular changes in the inclinations, but at present this is impossible.

In the Java Sea, between Borneo and Java, the tide is mainly diurnal; in fact according to the 'Eastern Archipelago Pilot,' part 3, the semi-diurnal tide is not appreciable on the north coast of Java till east of Surabaya, which is itself almost at the eastern end. On the south coast of Borneo the observations are not so numerous, but it seems clear that there also the tide is mainly or entirely diurnal. The tidal currents are described as weak. We can accordingly neglect the dissipation in the Java Sea. At its western end this sea is connected to the South China Sea by two rather wide straits, Carimata Strait and Gaspar Strait. In both of these the tide is mainly diurnal; in the latter, in fact, it is entirely so. At Pontianak, near the north-westerly point of Borneo, the tide is still diurnal; thus the part of the China Sea south of Pontianak and Singapore probably contributes little to the secular acceleration of the moon.

Apparently the main tidal stream from the China Sea strikes the Malay Peninsula somewhere near Cape Patani and spreads out from there; for on the coast north of this point the flood stream sets to the north, while south of it it sets to the south. The tide in this region is definitely semi-diurnal, though the heights of the two daily high waters may be unequal. The depths in the western part of the sea and in the Gulf of Siam are mostly about 30 fathoms, but there are many shoals around the coast where the depth is only a few fathoms, and it is therefore necessary to be very critical of the sites of observations of currents. The best results seem to be given at small islands with rapidly shelving sides, for the currents there are modified little by the form of the bottom and can be regarded as fairly typical of the general currents in the neighbourhood.

At the Anamba Islands (see map, fig. 1) the semi-diurnal tide appears to be usually much less than the diurnal one. The 'China Sea Pilot,' vol. 3, states that for a few days in each month, when the moon is near the Equator, there are two high tides in the day. It is easily seen that for two tides to occur in the day the amplitude of the semi-diurnal term must be at least a quarter of that of the diurnal term. It would not, however, vary much with the moon's declination, whereas that of the diurnal term vanishes when the moon is on the Equator; and the above fact shows that the semi-diurnal tide only attains this fraction of the diurnal tide when the latter is at its least. The true semi-diurnal tide at the Anamba Islands must therefore be insignificant. The same is evidently true of the currents, for the tidal streams take a day to run backwards and forwards.

In the Gulf of Siam also the tides are mainly diurnal. The oscillation in this is a forced one due to that in the South China Sea, and as the latter is diurnal so is that in the Gulf of Siam. Actually the only place where the semi-diurnal tide is considerable is Bangkok Harbour, at the head of the gulf. This tide seems to increase in relative importance towards Bangkok, for at Kamput on the eastern side and places in about the same latitude on the western side the tides are said to be very irregular, indicating the presence of some complicating influence. On the whole, therefore, it seems that the dissipation in the Gulf of Siam will not be underestimated if we assume that the semi-diurnal current reaches a maximum of one knot, this being one-third of the diurnal tidal current observed off Cape Patani, at all places north of the parallel of 11° N. The area of this region is about 70,000 sq. km., giving a maximum dissipation of $2 \times 10^{17}$ ergs per second.

An estimate may be made of the dissipation of the energy of the diurnal tide in the same regions. At Pontianak there is a diurnal current of two knots when running strongest; and a similar velocity is recorded at the Burong Islands some distance to the north. As the depth of the passage between Borneo and Sumatra is fairly uniform, from 10 to 20 fathoms, these measures are probably typical of the whole. Quantitative estimates of the currents in the Gulf of Siam are few, and a reasonable guess at them would be difficult. The depth is mostly about 30 fathoms, but there are a deeper area in the middle and many shoals about the margins. As the current at the mouth of the gulf is across it, the tide in the gulf can arise only from the reflection of this by the Malay coast, so that the general set of the current is across the gulf, and considerable magnification in the gulf is unlikely. Our estimate of the dissipation will probably be of the correct order of magnitude, if we suppose that everywhere west of a line joining Cape Datu to Cambodia Point the maximum current is two knots. The area of this region is $8 \times 10^5$ sq. km.; thus the maximum dissipation is $1 \cdot 7 \times 10^{19}$ ergs per second. The velocity is proportional to the sine of the hour angle of the moon increased by a constant; and as the dissipation is proportional to the cube of the velocity, the average dissipation is obtained by

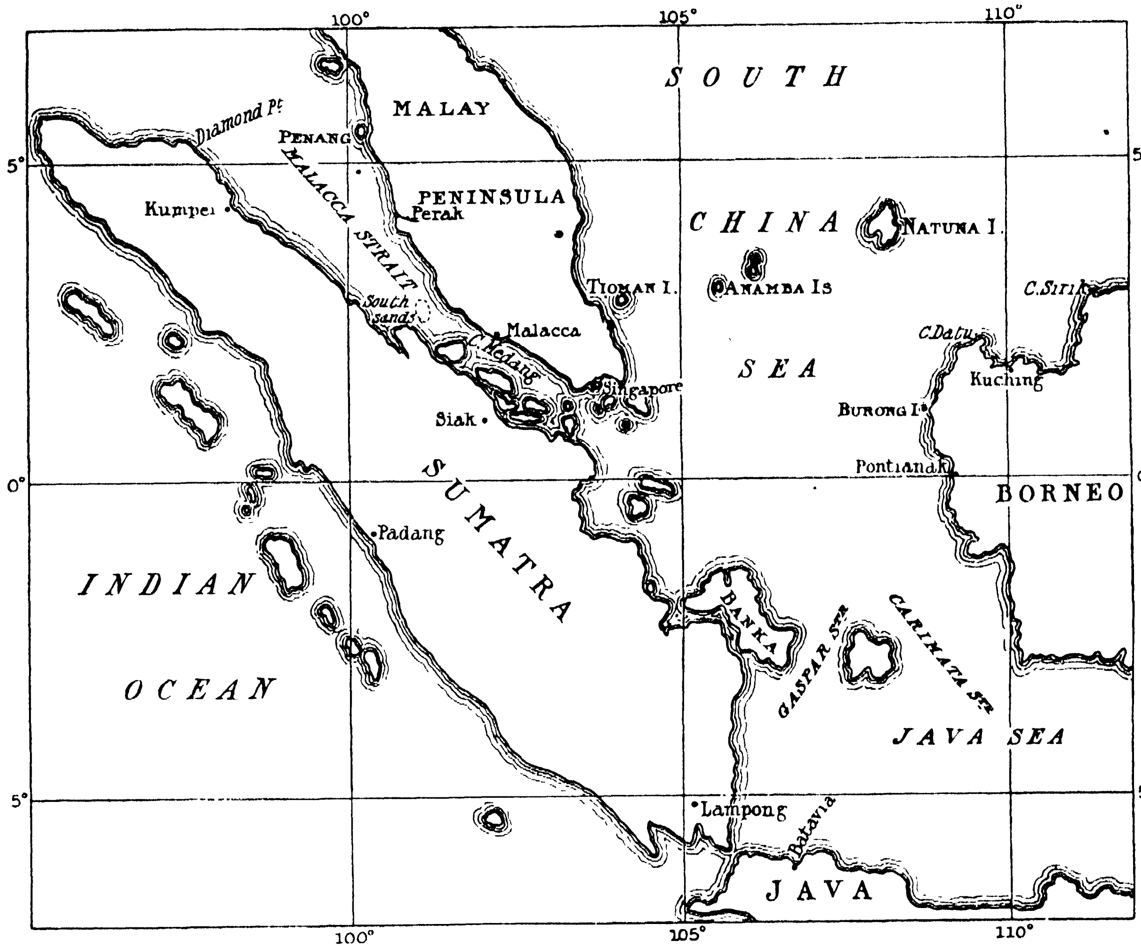


Fig. 1.

multiplying the maximum by the average arithmetical value of $\sin^3 \theta$ taken over a period, which is $4/3\pi$. It is not clear what declination of the moon the recorded currents refer to: if they refer to the maximum, the average for the month will be found by multiplying again by $4/3\pi$. Thus the average dissipation of energy in the diurnal tide in the western part of the China Sea is of the order of $3 \times 10^{18}$ ergs per second.

We next proceed to examine the dissipation in the main part of the South China Sea, between the north-west coast of Borneo and the mainland. The currents are now semi-diurnal. Near Cape Sirik the flood runs for four hours and the ebb for eight hours, so that a considerable diurnal component exists, but not sufficient to preponderate over the semi-diurnal motion. The velocities here are from two to three knots. At Bruni the observations are very much interfered with by shoals and narrows, but in the offing the currents seem to be about two to three knots. About Tega, however, among shoals the velocities recorded are only 1·3 and 0·8 knots. The contours of the sea floor run roughly parallel to the coast, so that these currents may persist for some distance out to sea. Off the north end of Borneo the sea rapidly deepens, and in accordance with this there is scarcely any tidal current at Ulugan Bay, in Palawan. On the Asiatic side the tide is diurnal at Camran and Tourane, but the currents are weak.

A few shoals and islands in the middle of the sea have been made the localities of observations. At Rifleman Bank and Spratly Island there is only one tide in the day, and at the neighbouring island of Amboyna it is said that near neaps the stream reaches 1·4 knots. It is therefore clear that the semi-diurnal current of the Borneo coast does not extend half way across the sea, and its true extent is very doubtful. In the Gulf of Tongking also the currents appear to be diurnal. Thus in the whole of the South China Sea and its extensions there seems to be little semi-diurnal tide and little contribution to the secular acceleration of the moon, though there is a dissipation of the energy of the diurnal tide that may have a notable secular effect on the obliquity of the ecliptic.

### 2. *The Yellow Sea.*

This is a gulf about the size of Ireland, lying between Korea and the coast of China, and extending about as far south as the mouth of the Yang-tse-Kiang. It becomes very narrow where the Shan-tung peninsula projects into it, and north of this it forms the Gulfs of Pe-Chili and Liau-tung. Most of it is shallow, the depths in the main part of it being mostly about 30 fathoms, and those in the northern part about 15 fathoms. Around the shore the water is shallower, and in many places there are crowds of shoals. The data are obtained from the 'China Sea Pilot,' vol. 5.

The tidal phenomena are extremely complex. At the south end of the peninsula of Korea high water (full and change) is at about 11h., Greenwich time. As we advance up the Korean coast it occurs later and later, being practically 12 hours later at Port Arthur. On the opposite side of the Strait of Pe-Chili it is slightly earlier than at Port Arthur; then on the way round the Shan-tung peninsula and out of the sea again the tide again becomes steadily later. It therefore looks as if the tide enters up the coast of Korea, gradually passes up the sea, losing energy all the way, and a reflected wave from the Pe-Chili strait emerges down the Chinese coast. The tides and the currents on the Korean side are noticeably stronger than those on the Chinese side, and it does not seem likely that this is due wholly, or even largely, to the shoals on the former coast, for at islands in deep water, such as the Mackau, Myangoru, and Bate groups, velocities of 3 to 5 knots are recorded, while in shallower water the velocities are not usually greater than these, though local strong streams exist. Apparently the effects of friction are great enough to counterbalance those of the diminution in depth. They are also seen in another respect. Among the islands of the Korean archipelago the tidal stream sets west from four hours before till two hours after high water, whereas in most places elsewhere there is little or no current at high water. Thus work is continually being done on the sea, and the energy entering is dissipated in it. An effect of the approximate agreement in phase between the tide and the velocity may be utilized to give an estimate of the currents and tide height in the entrance to the sea, far from the nearest land, and hence of the amount of energy entering. In any motion in a channel where, on

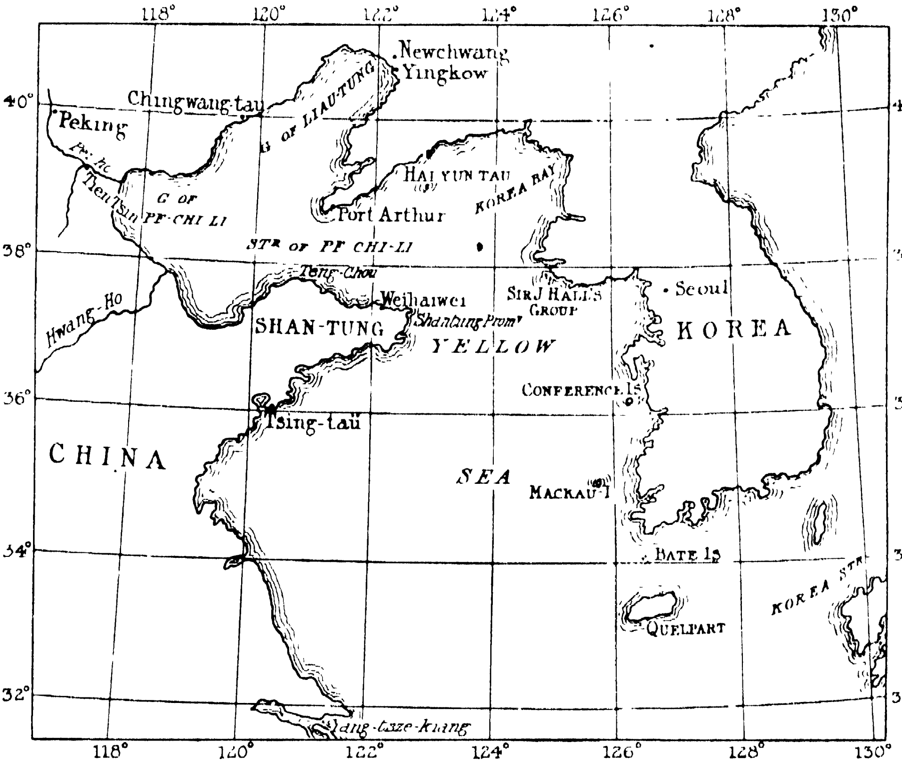


Fig. 2.

account of its narrowness or for some other reason, there is little transverse motion, the velocity along the channel is related to the pressure gradient across it according to the equation

$$2\omega\rho u = -\partial p/\partial y,$$

where $\rho$ is the density of the water,

$\omega$ is the component of the earth's angular velocity of rotation about the vertical at the point considered,

$p$ is the pressure, and

$y$ is the distance measured across the channel.

In the Yellow Sea the entrance is not narrow, but there seems reason to believe that the velocity across it is small, which is all that is required for the truth of the above equation. If now $g$ denote the intensity of gravity, and $\eta$ the elevation of the surface of the water above its mean position, then at any fixed point, at whatever depth, the variation of $p$ is equal to that of $g\rho\eta$. Again, if $\Omega$ be the earth's angular velocity, and $\lambda$ the latitude of the place,

$$\omega = \Omega \sin \lambda,$$

and we have on putting $\lambda = 35^\circ$; $\Omega = 7{\cdot}3 \times 10^{-5}/1$ sec.; $g = 981$ cm./sec.$^2$;

$$u = -1{\cdot}17 \times 10^7 \frac{\partial \eta}{\partial y}$$

where C.G.S. units must now be used.

On the coast of Korea the tide has an amplitude of about 10 feet, or 300 cm. The velocity of the inward current is about 4 knots, or 200 cm./sec. Now suppose if possible that the current remained constant right into the middle of the entrance; then the above formula shows that at a distance of 176 km. from the side there would be little vertical movement of the surface, and further away still a huge tide with an amplitude of some 30 feet would exist. It is not reasonable that the tide in the middle should be greater than that at the side, though it may easily be smaller. The alternative hypothesis is therefore that the current decreases as we approach the middle, and is very small over most of the sea. This will be adopted in the forthcoming discussion. We shall suppose that the current at distance $y$ from the coast is in the same phase as that at the coast, and is a linear function of $y$. Then put

$$u = (200 - ky) \cos \gamma t.$$

At the shore $\eta$ is equal to $300 \cos (\gamma t - \alpha)$, where $\alpha$ is the difference in phase between the tide height and the current strength. For the semi-durnal tide it is twice the angle moved through by the moon relatively to the earth in one hour, or 29 degrees. In general

$$\eta = 300 \cos (\gamma t - \alpha) - 8{\cdot}6 \times 10^{-8} \cos \gamma t \,(200y - \tfrac{1}{2}ky^2).$$

The amplitude of $\eta$ reaches a minimum where the coefficient of $\cos \gamma t$ vanishes. We shall have the least possible tide in the middle, thus satisfying our earlier assumptions, if we assume that this minimum is reached at the point where the velocity vanishes. In this way we shall underestimate the dissipation, but not by any great amount. The value of $k$ that makes this coincidence possible is $6 \cdot 6 \times 10^{-6}/1$ cm., so that the current becomes zero 300 km. from the shore, practically in the middle of the entrance.

Taylor shows (*loc. cit.*, equation 15) that the average amount of energy crossing any line is the average over a period of $g\rho \int \mathrm{D}\eta u \, dy$ in my notation, where D is the depth of the water, in this case about 30 fathoms over most of the region in which the velocity is greatest, with a range in all of from 20 to 45 fathoms. We find easily

$$\eta = 147 \sin \gamma t - 4 \cdot 3 \times 10^{-8} \, k \, (y - 200/k)^2 \cos \gamma t,$$

and the average flux of energy across a parallel of latitude is found to be $1 \cdot 06 \times 10^{18}$ ergs/sec.

The northward flux on the Chinese side is more difficult to determine, as the direction of the currents is variable. Two phenomena are intermingled here. The issuing tide from the Yellow Sea comes down this coast, but there is also a definite tidal wave that travels into and out of the large bend in the coast whose extremities are Shan-tung promontory and Shanghai. This is shown by the fact that along the northern part of this bend, on which Tsing-tao stands, the current flows northwards while the tide is ebbing both along this coast and in the northern part of the Yellow Sea. Thus in this bay the main tide is the local tide of the bay itself and not the general tide of the Yellow Sea. The currents produced by these tides are rotary, probably an effect of the earth's rotation ; and it seems that the northward component of the velocity is small. Further, it is probably nearly in a phase at right angles to the tide, as great divergences from this relation can be produced only by great dissipations and accordingly by great velocities in the vicinity. Hence for both reasons we infer that the northward flux of energy along the Chinese coast is small in comparison with that on the Korean side.

We also need to know the work done on the water by the moon. If $\eta'$ is the height of the equilibrium tide, the work done by the moon in a period is $\iint g\eta' \frac{d\eta}{dt} \, dt \, d\mathrm{S}$, where $d\mathrm{S}$ is the element of horizontal surface and the integrals are to be taken, in the one case over a period, and in the other over the area considered. If the phase of $\eta$ is $\beta$ in advance of that of $\eta'$, and the amplitudes be $h$ and $h'$, the average rate of doing work is $-\frac{1}{2} g \int h h' \gamma \sin \beta \, d\mathrm{S}$. In the present case $h'$ naturally varies little ; $h$ decreases fairly steadily as we travel from the entrance to Port Arthur, where it has about half of its value at the entrance. On the other hand $\beta$ varies a great deal. The longitude being about 120° E., the moon crosses the meridian at full and change at 3h. 43m. It is high water in Shoan harbour, at the southern end of Korea, at 10h. 33m., so that $\beta$ is here $-199°$, or more conveniently $+161°$. The tide lags more and more on the way up the sea, and at Port Arthur it has lost practically a whole period, the high water at full and change occurring there at 11h. 7m., making $\beta = -216°$. Positive elements will be added to the integral by the places where $\beta$ is from 161° to 0°, negative elements where $\beta$ is from 0° to $-180°$, and positive again where it is from $-180°$ to $-216°$. These values are however weighted according to the values of $h$ and according to the extent of the areas for which they are correct. Now $\beta$ is zero near the Conference Islands, only about 0·4 of the way to the narrowest part, so that on this account the area for which it is positive would appear to be less than that for which it is negative. The sea becomes much narrower farther north, however, which must reduce the ratio of the weights somewhat, and the tides are only about two-thirds of the height. Thus it seems that the weights to be attached to the positive and negative values of $\sin\beta$ are nearly equal, and its average over the sea is unlikely to be more than 0·2. The area of the sea as far as Port Arthur is about 300,000 sq. km.; taking the average of $h$ as 240 cm., and that of $h'$ as 20 cm., we find the energy imparted by the moon to be not greater in absolute magnitude than $2 \times 10^{17}$ ergs per second on an average. That entering up the Korean coast is far greater.

In the discussion of the work done by the moon the Gulfs of Pe-Chili and Liau-tung have been ignored. There are two reasons for this: their united area is about a third of that of the main part of the sea, and the tides recorded at the sides are also about a third of those on the Korean coast. It seems to me, however, that these two gulfs afford an example of a special type of tidal problem different from any previously discussed. For, let us suppose if possible that the recorded amplitudes, of the order of 90 cm., were typical of the whole area. The average depth is about 2000 cm., and a tidal wave in water of such a depth would give rise to a current of maximum velocity about $h\,(g/\mathrm{D})^{\frac{1}{2}}$, which in this case is 65 cm./sec. The corresponding dissipation would be 550 ergs per square centimetre per second. On the other hand the average energy present is about $\frac{1}{2}g\rho h^2$, or $4 \times 10^6$ ergs per square centimetre. Thus if the above assumption were correct the whole energy of the tide would be dissipated in about 7000 seconds, or two hours. This is absurd, and we must suppose, in order to avoid the result that energy is dissipated faster than it enters the region, that the tides in the greater part of the gulfs are much less than the recorded ones. Their height may be only a few inches; while the recorded heights are the result of great magnification in the very shallow water around the edge. Accordingly the work done by the moon on this region may be neglected. The whole work done on the Yellow Sea by the moon is therefore small in comparison with the energy entering with the tide, and even its sign is uncertain. Thus the average dissipation in the whole of the sea is not very different from $1{\cdot}1 \times 10^{18}$ ergs per second.

An alternative estimate may be deduced directly from the formula for the dissipation, with a suitable hypothesis on the distribution of velocity. At the entrance, when the current is flowing strongest, the dissipation in a strip a centimetre wide across it is $0 \cdot 002\rho\int u^3 dy$, and if the above distribution of velocity is correct this is $1 \cdot 2 \times 10^{11}$ ergs per second. Farther north the velocity is not so great; in fact, around the Shan-tung promontory it does not exceed one knot, and at the islands in Korea Bay opposite it is usually about two knots. Further, the width is here less than 300 km., so that on this account also the dissipation for a given velocity must be less; the amount in a strip a centimetre wide running east and west is therefore probably not more than one-sixteenth of that in a similar strip near the entrance. Even south of the narrow part off Shan-tung the currents appear to be slower than near the entrance; for about half the distance the velocity is about two knots, rising again to $3\frac{3}{4}$ knots at the Sir James Hall group, in the narrow region. If a proportional reduction takes place at all other distances from land, we must suppose that the above estimate of the dissipation per unit length is correct for the first 200 km., that the amount for the next 230 km. is an eighth of this, and that for the remaining 220 km., corresponding to Korea Bay, is a sixteenth. The Gulfs of Pe-Chili and Liau-tung may be ignored. The total dissipation, if the maximum velocity at every place occurred at the same time, would therefore be $(2 \times 10^7 + \frac{1}{8} \times 2 \cdot 3 \times 10^7 + \frac{1}{16} \times 2 \cdot 2 \times 10^7)\ 1 \cdot 2 \times 10^{11}$ ergs per second, or $2 \cdot 8 \times 10^{18}$ ergs per second. The average for each place is $4/3\pi$ of the maximum there, so that the average for the whole sea is $4/3\pi$ of this maximum, or $1 \cdot 2 \times 10^{18}$ ergs per second, which agrees with the previous estimate much more closely than the data would have led us to expect.

### 3. *The Sea of Japan.*

The Sea of Japan is an oval basin bounded on the eastern side by Japan and Sakhalin. It seems clear that the dissipation is small, for both the tide height and the current are small. Even in the comparatively narrow and shallow Korea Strait, through which the tide enters at the south end, the current only attains a speed of 1 or 2 knots; and when this opens into the sea the width suddenly increases to 900 km. and the depth to 400 fathoms. The currents in most of the sea must therefore be insignificant. They are appreciable at the gaps between the Japanese islands and in part of the narrow Gulf of Tartary, but the area affected is small and the currents only moderate (1 to 3 knots at most) so that the dissipation is small.

### 4. *The Sea of Okhotsk.*

In its essential features this resembles the Sea of Japan. The only shallow parts of it are narrow strips around the coast, while the tide enters through the shallow water of the straits between the Kurile Islands. As the tide in the sea depends on the supply of water to maintain it, the restriction on it imposed by the shallowness of the entrances causes the currents to be small. In the Gulfs of Ghijinsk and Penjinsk, in the north-east corner, the depth diminishes considerably, and the currents increase to $1\frac{1}{2}$ to 2 knots. Data are very scanty, but the area affected appears to be about 70,000 sq. km., leading to a dissipation when the currents are strongest of from $6 \times 10^{17}$ to $1 \cdot 4 \times 10^{18}$ ergs per second. If we adopt the mean of these as giving roughly the actual dissipation, and apply the factor $4/3\pi$, we find the average dissipation in these gulfs to be about $4 \times 10^{17}$ ergs per second. There is probably no important dissipation elsewhere in the Sea of Okhotsk.

### 5. *The Bering Sea.*

In the extreme north of the Pacific, between Siberia and Alaska, a chain of small islands, the Aleutian Islands, extends all the way across. The region north of these has the shape of a quadrant and forms the Bering Sea. Between the islands the depth is great, and the tide of the Pacific seems to enter almost unhindered. Since the depth of more than half of the sea, mostly on the Alaskan side, is less than 40 fathoms, large currents are produced, especially in the three chief bays—the Gulf of Anadir, Norton Sound, and Bristol Bay. The dissipation must therefore be very

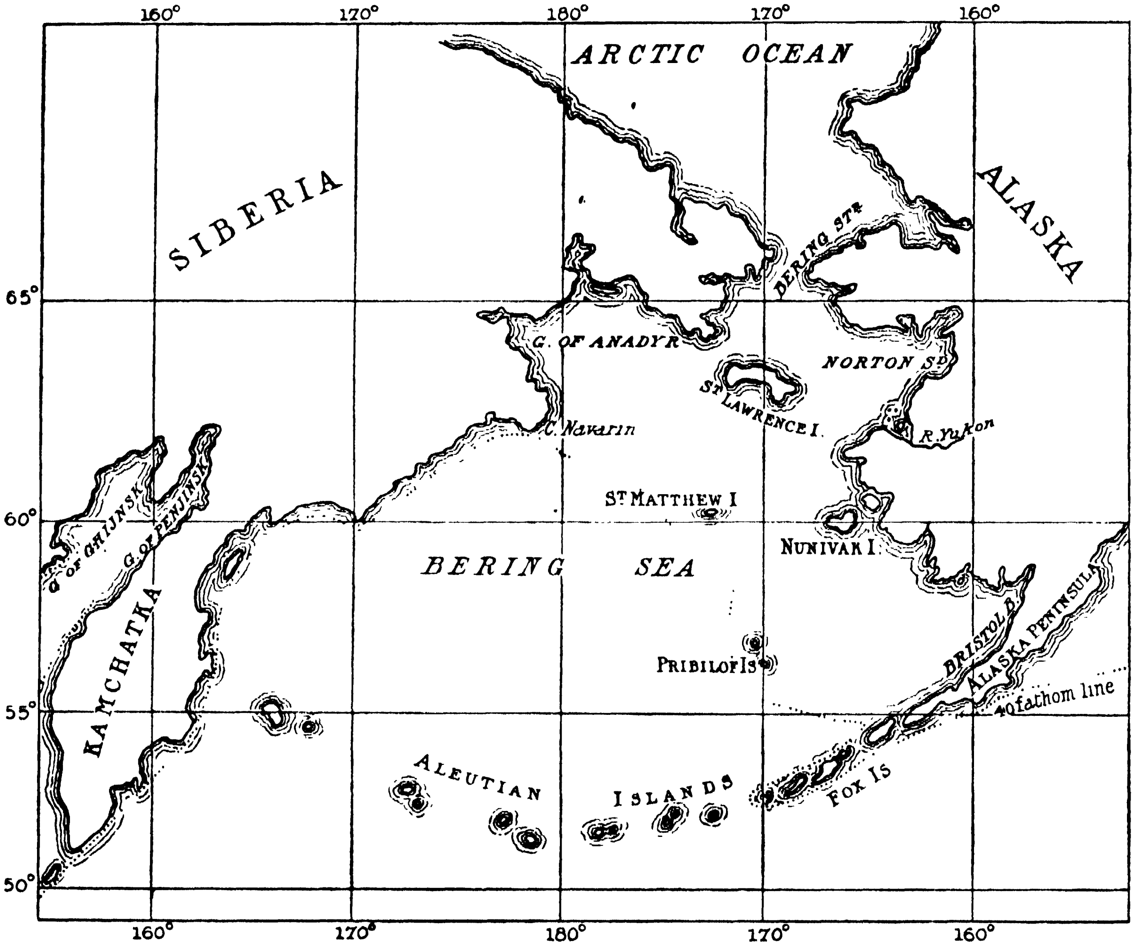


Fig. 3.

great, but a reasonably accurate estimate of it is difficult to make on account of the form of the shallow portion, which has no narrow place that can be called an entrance. It is best to treat the main part of the sea and the bays separately.

In the south of the sea it is stated that the maximum rate of the water, when clear of the passes between the Aleutian Islands, is usually about $2\frac{1}{2}$ knots when the depth is less than 100 fathoms. In the region satisfying these conditions the depth is in most places about 80 fathoms, so that the current farther north, where the depth is often only 20 or 30 fathoms, may exceed this. On the other hand there seems to be little semi-diurnal tide in the extreme north. In Norton Sound the tide is diurnal, presumably because the waves from the south and from Bering Strait neutralize each other. At St. Lawrence Islands, near the entrance to the strait, the tide is only about a foot in height, confirming this suggestion. Farther south, however, the tidal wave from the Arctic must spread out and become inappreciable. At Pribilof Islands, 500 km. from the nearest land and surrounded by water 50 fathoms deep, the current reaches $2\frac{1}{2}$ knots. At St. Matthew Island, about midway between these and St. Lawrence Islands, in water 30 to 40 fathoms deep, and nearly as far from land, the current still reaches $2\frac{1}{2}$ knots. There are no other data given for islands far from shore, and it seems that we shall not be overestimating the dissipation if we take the maximum current to be $2\frac{1}{2}$ knots all over the shallow region bounded on the south by the Fox Islands and extending north till half-way between St. Matthew and St. Lawrence Islands. The size of this is 1,000 by 700 km., or $7 \times 10^{5}$ sq. km. The maximum dissipation is therefore $2{\cdot}74 \times 10^{12} \times 7 \times 10^{5} \times (2{\cdot}5)^{3}$, or $3 \times 10^{19}$ ergs per second, and the mean dissipation $1{\cdot}2 \times 10^{19}$ ergs per second.

In Bristol Bay the average velocity seems to be about 3 knots, though the observations are few. The corresponding dissipation is about $1{\cdot}5 \times 10^{18}$ ergs per second. In Norton Sound the dissipation is probably small, for it is mostly north of St. Lawrence Island, and the tide is diurnal. The Gulf of Anadir probably contributes about as much as Bristol Bay; for though its area is twice as great, its more northerly situation must reduce the current somewhat. In all, then, the average rate of dissipation in Bering Sea is about $1{\cdot}5 \times 10^{19}$ ergs per second. This estimate is of course subject to considerable error, for it depends wholly on a few observations, which may not give quite a fair sample of the whole of the sea. The depths around the localities considered are fairly typical of the sea as a whole, so that great error on this ground is not to be anticipated; but errors in observing the velocities may be greater, and both kinds of error are magnified in importance by the fact that the velocity must be cubed when the dissipation is calculated, so that if the true mean velocity were only 2 knots instead of $2\frac{1}{2}$ knots the dissipation would be almost halved. It does not appear that the velocity increases much towards the coast; in fact the velocities near the Alaskan coast seem to be rather smaller than those near the islands. Thus an underestimate on this ground is not probable.

### 6. *Malacca Strait.*

This is a narrow triangular area, about 800 km. in length, separating Sumatra from the Malay Peninsula. The tide of the Bay of Bengal enters at the north-west end, and gradually increases in height as it advances along the strait towards Singapore. At the south end, however, the part of the tide that has not been reflected or dissipated on the way through the strait is overwhelmed by the diurnal tide of the South China Sea. Ample observations of the tides and tidal currents on both sides are available. The currents as far south as Cape Medang (nearly due west of Malacca) seem to reach maxima of 1½ to 3 knots, the average amplitude at springs being practically 2 knots. The area of this region is 100,000 sq. km., and the dissipation is accordingly found to be about $9 \times 10^{17}$ ergs per second on an average.

An alternative determination can be made by finding the rate of inflow of energy. At Kumpei, on the Sumatran side and near the north end of the strait, it is high water, full and change, at noon, and the amplitude at springs is 120 cm. The flood tide outside the bar, in water about 20 fathoms deep, sets south-east from three hours before high water till three hours after it, so that it reaches its maximum speed of 1½ knots at high water. At Penang, near the opposite shore, it is high water at 0h. 21m. and the current reaches its maximum velocity of 2½ knots an hour before high water. As the strait is everywhere narrow in comparison to its length, and as these observations do not seem to have been taken on shoals, they are probably representative of that part of the strait. The amplitude of the tide at Penang is 100 cm. We can therefore take the average height of the tide along the section from Kumpei to Penang to be 110 cm., and the average current when flowing strongest to be 2 knots, reaching its maximum half an hour before high water. The average depth is about 30 fathoms. A modification must be made in the previous procedure to allow for the fact that the current flows along the strait, which is not quite at right angles to the line of the section; therefore, in finding the energy crossing the section, we must take for the length of the section, not the distance from Kumpei to Penang, but the projection of this on a line perpendicular to the strait, which is 230 km. The flux of energy is hence found to be, on an average, $7 \times 10^{17}$ ergs per second. We also require the amount of this energy that emerges through the narrow part of the strait. Off Cape Medang, which marks the narrowest point of the strait away from the immediate neighbourhood of Singapore, the amplitude of the tide is 120 cm., high water occurring at 6h. 30m. The tidal current flows at an average speed of about 2½ knots. There is no record of the tidal phenomena just opposite, but in Malacca Road it is high water at 7h. 30m., with an amplitude of 165 cm.; the current there reaches its maximum of 2 knots an hour before high water. At the eastern end of South Sands, which lies near the Malay side, north-west of Cape Medang, it is high water about 6h. 0m., and the tidal stream has a maximum speed of 1½ knots an hour before high water. The width of the channel at Cape Medang is

36 km., and the depth 20 fathoms. The average flux of energy eastward past it is found to be $1 \cdot 0 \times 10^{17}$ ergs per second. Thus the average excess of the inflowing energy over the issuing energy is $6 \times 10^{17}$ ergs per second. The work done by the moon is insignificant, for it crosses the meridian at full and change at 2h. 30m., which is nearly the average time of high water. Thus all the excess of energy just found is dissipated in the strait.

The area of the strait between the Kumpei-Penang section and the Medang section is 56,000 sq. km. If the average current in this had an amplitude of 2 knots the dissipation would be $5 \times 10^{17}$ ergs per second on an average, in striking agreement with the estimate from the flux of energy, though the latter is more reliable. If in the final estimate the region north of the Kumpei-Penang section is to be included, we must add a fraction to the total to allow for it, making probably between $8 \times 10^{17}$ and $12 \times 10^{17}$ ergs per second in all.

In the part of the strait east of Medang there are few records of the currents, but the dissipation is probably small. In any case it could not exceed the $10^{17}$ ergs per second that pass Medang, and is probably less than this. The total dissipation in the Strait of Malacca is therefore $1 \cdot 1 \times 10^{18}$ ergs per second, subject to an uncertainty of a fifth of its amount.

## AUSTRALIAN WATERS.

Australia is surrounded by a belt of water less than 100 fathoms in depth ; the width of this ranges from 10 to 200 miles, except at the Gulf of Carpentaria, where it extends right across to New Guinea. The tide in this neighbourhood is diurnal, like that in the South China Sea to the north of it. The contribution to the secular acceleration of the moon is accordingly very small. The tidal streams do not exceed 1 knot, and as the area is much less than that of the South China Sea the dissipation in the diurnal tide cannot be comparable with that already found for the larger sea.

## AFRICAN WATERS.

### *The Mozambique Channel.*

The channel between Madagascar and the mainland is mostly about 500 fathoms deep or more. There are few records of tidal currents in it; in fact the only record given in the 'African Pilot,' part 3, appears to be based on the statement of a single observer, that the tidal streams in the channel are comparable with the permanent current driven by the trade winds, which flows at about 2 knots. This cannot however be uniform all over the channel, for the following reason. The height of the tide along the African coast is about 12 feet, which is as usual measured relative to low water at ordinary springs, so that the vertical amplitude of the tide is 180 cm. Now the ordinary theory of tides in channels shows that the maximum velocity is of order $h\,(g/\mathrm{D})^{\frac{1}{2}}$; in a simple wave in a uniform channel it is exactly this. Taking the depth to be 90,000 cm., this makes the maximum velocity 19 cm./sec., or rather more than a third of a knot. Accordingly the currents with velocities of a knot or more must be confined to narrow coastal strips, and the dissipation is therefore small.

The only other partially enclosed regions around Africa are the Gulf of Aden and the Red Sea. The former is deep in the middle with narrow strips of shallow water on the margins, like the Mozambique Channel, and therefore the dissipation is small. The Red Sea is shallower, but can have no important currents, since the inlet at Aden is so narrow. The Mediterranean has already been dealt with. Thus the dissipation around the coasts of Africa is negligible.

## NORTH AMERICAN WATERS.

There are many partially enclosed bodies of water around North America, the chief of which are the Gulfs of Mexico and California, the Bay of Fundy, the Gulf of St. Lawrence, and the numerous straits and bays of the North-west Passage. Of these the Gulf of Mexico may be ruled out at once, for it is very deep and a large fraction of its entrance is blocked by Cuba. The Gulf of California is still deeper; and therefore the currents in these cannot be notable except in restricted localities.

### 1. *The Bay of Fundy.*

This bay requires to be considered separately in spite of its small size, for it is famous for possessing the largest tides in the world. It is fairly shallow, and the tides are much magnified in height by the diminution in both depth and width towards the head of the Bay. The currents are apparently not so great as would be expected from the height of the tides. The entrance is through the Grand Manan Channel, named after an island in it. The average current in the channel reaches about 1·8 knots, and that near St. John, half-way up the bay, reaches 1·7 knots. The area of the bay is 12,000 sq. km., so that the average dissipation for a maximum velocity of 1·8 knots all over would be $7{\cdot}7 \times 10^{16}$ ergs per second. It is likely that the currents farther up the bay are stronger, so that this must be regarded as a lower limit.

An alternative estimate may be obtained from the inflow of energy. The rise of the tide in Grand Manan Channel is at most places about 20 or 22 feet at springs, above low water ordinary springs. The amplitude is therefore about 320 cm. The current has an amplitude of 1·8 knots, and the depth of the channel is about 9000 cm. The average time of the turn of the current at three places near the south side of the channel (those numbered 14, 16 and 17, in the 'Nova Scotia and Bay of Fundy Pilot,' page 22) is 35 minutes after high water at St. John. This high water at full and change occurs at 11h. 21m., while at l'Etang, on the north side of the channe it is at 11h. 18m., and the mean of the times at Westport, Petit Passage, and Digby Gut, which are nearly opposite, is 10h. 48m. Thus the mean time of high water in the channel must be about 11h. 3m., so that the current turns 53 minutes after the tide. The phase difference is therefore 25 degrees. The width of the channel is 83 km. Applying equation 15 of TAYLOR's paper, we find that the average rate at which energy enters is $4 \cdot 7 \times 10^{17}$ ergs per second.

The average rate at which the moon does work on the bay is $-\frac{1}{2}g \int hh'\gamma \sin \beta dS$, as was found for the Yellow Sea. In this case the latitude is 45° north, so that $h'$ is 18 cm. In the lower half of the bay the amplitude of the tide is not greater than 12 feet, but in the upper half it rapidly increases, till in Minas Basin it reaches 25 feet and in Chignecto Bay 23 feet. The time of high water in the bay ranges from 11h. 3m. to 11h. 50m. The later value corresponds to the upper part, where the amplitude is greatest; but as this part is also the narrowest, the two times must receive about equal weights in finding the average. We therefore take the average time of high water to be 11h. 30m. The average amplitude of the tide is about 18 feet, or 540 cm. The longitude of the bay is 66° west, so that the moon crosses the meridian at full and change at 4h. 33m. The time of high water is more than 6h. 12m. later than this, so that the tide is falling when the moon is exerting its greatest upward pull, and the work done by the moon is therefore negative. The interval between transit and low water is 45 minutes, so that $\beta = 22^\circ$. The area of the bay is $1 \cdot 2 \times 10^{14}$ sq. cm. The work done by the moon is therefore $-3 \times 10^{16}$ ergs per second. The total dissipation in the bay is $4 \cdot 4 \times 40^{17}$ ergs per second.

This estimate is six times as great as the earlier one based on the currents alone. It is much the more reliable, for the first depended on the assumption that the currents were equally great all the way up the bay, whereas actually they increase very much towards the head. Velocities up to 9 knots are recorded in Minas Basin, though the area in which these occur must be very restricted. The most serious source of error in the second estimate is the phase difference, for this is only an hour and would be affected to a considerable extent by an error in the determination of the time when the current turns. The observed time of turn does not vary much from place to place, however, and it does not seem likely that the estimate is wrong by more than a quarter of its amount. The second estimate will therefore be adopted. It will be noticed that it is rather less than the dissipation in the Irish Sea.

The Gulf of St. Lawrence gives very little dissipation. The narrow entrance through Cabot Strait prevents the tides from being considerable except in the estuary of the river itself and in Belle Island Strait, which separates Newfoundland from Labrador.

### 2. *The North-west Passage.*

The channel from the Atlantic to the Arctic between Canada and Greenland is blocked by a large number of islands of varying sizes, between which are narrow and shallow straits. The dissipation in several of these can be estimated from data in the 'Arctic Pilot,' vol. 3.

The chief of these channels is Davis Strait, with Baffin Bay to the north of it, which lies between Baffin Land and Greenland. It is about 1600 km. in length. The only tidal velocities recorded in it, except in fjords, are near Holstenborg, in Greenland, where the currents in the offing are said to reach a speed of two knots. This cannot, however, be general, for there is a shallow region off Holstenborg, some 150 km. long and 60 km. wide, with a depth of about 23 fathoms. Most of the strait is about 100 fathoms deep. This region is therefore a place where the main current of the strait is magnified by the form of the bottom, and there is no reason to believe that the current in the deep water is greater than half a knot, which is the observed velocity off the coast of Labrador. The dissipation in Davis Strait is therefore not great.

At the northern extremity of the strait there are several narrow passages into the Arctic. The dissipation in these must be important. In Smith Sound and Kennedy Channel, for instance, which separate the north-west coast of Greenland from Ellesmere Land, the current is said to be "nearer two figures than one." These straits are small in area, but if such currents exist over much of their extent we must take them into account. The data available at present are unfortunately too meagre.

The south end of Davis Strait is connected to Hudson Bay by Hudson Strait. The currents in this are described as "great enough to be dangerous," especially at the east end; but the recorded currents, even in the middle of the strait, are only about three-quarters of a knot. This makes the average dissipation about $5 \times 10^{16}$ ergs per second. The danger arises mostly from drifting ice.

In the entrance to Hudson Bay the velocity increases to one and a half knots. The area over which this is true is about $3 \cdot 8 \times 10^{14}$ sq. cm., making the average dissipation $1 \cdot 5 \times 10^{17}$ ergs per second.

In Hudson Bay itself the currents are probably very small. Considerable velocities are recorded at Port Churchill, but there are no records in the middle of the bay. The depth in the middle is about 50 fathoms, which is about the same as at the entrance. The entering current must therefore spread out in the bay and undergo great diminution in strength. Near Port Churchill the depth is only 19 fathoms or less, so that the current there must be a local current magnified. The dissipation in Hudson Bay must therefore be small.

In Fox Strait, which runs northwards from the entrance to Hudson Bay, the depth is less, about 20 fathoms, and the current reaches one and a half knots. The area of this channel is $2 \times 10^{15}$ sq. cm., and the appropriate average dissipation is about $1 \cdot 4 \times 10^{18}$ ergs per second. The remaining straits of the North-west Passage probably do not contribute nearly so much to the dissipation, for the energy of the entering wave must be mostly dissipated in the channels already dealt with, and partly through this and partly on account of the obstructive effect of the islands it is not likely that the straits farther north-west are very important, though this cannot be regarded as certain. The dissipation in the whole of the North-west Passage is thus about $1 \cdot 6 \times 10^{18}$ ergs per second on an average. Adding this to the amount found for the Bay of Fundy, we have for the whole of North America a total of $2 \times 10^{18}$ ergs per second on an average.

## SUMMARY.

The mean rates of dissipation in the lunar semi-diurnal tide found in the foregoing investigation are as follows:—

| | Ergs per Second. |
|---|---|
| European waters | $2 \cdot 4 \times 10^{18}$ |
| Asiatic waters: | |
| South China Sea | Small. |
| Yellow Sea | $1 \cdot 1 \times 10^{18}$ |
| Sea of Okhotsk | $0 \cdot 4 \times 10^{18}$ |
| Bering Sea | $15 \cdot 0 \times 10^{18}$ |
| Malacca Strait | $1 \cdot 1 \times 10^{18}$ |
| North American Waters: | |
| Bay of Fundy | $0 \cdot 4 \times 10^{18}$ |
| North-west Passage | $1 \cdot 6 \times 10^{18}$ |

The total thus accounted for is $2 \cdot 2 \times 10^{19}$ ergs per second. I have shown in a previous paper that the dissipation required to account for the secular acceleration of the moon (which amounts to $9''$ per century per century) is about $1 \cdot 4 \times 10^{19}$ ergs per second, so that it seems as if there is more dissipation than is required. If this was so it would be necessary to seek for a cause that could produce an appreciable secular retardation of the moon, and none such is known. A scrutiny of the results so far obtained is therefore desirable, with a view to finding out whether any of them have been over-estimated. One cause of such an over-estimate is easily seen. The data used for the Irish Sea, the English Channel, Malacca Strait and the Bay of Fundy refer definitely to spring tides alone when the currents are at a maximum. The height of the tide adopted in the calculation for the Yellow Sea was also that of the spring tide. In the other cases it is not stated whether the currents have average or spring values, but if they were determined at springs the requisite reduction is at once obtained. The theoretical ratio of the heights of the lunar and solar tides is $2 \cdot 3$ when inertial and frictional effects are neglected. This ratio is probably nearly correct in mid-ocean, for the periods of the two tides are not very different, so that inertia will affect their amplitudes in the same ratio. In shallow areas, however, the frictional force is not proportional to the velocity but to its square, and accordingly friction has more relative effect in reducing the tides when they are great than when they are smaller. The ratio of the solar to the lunar tide, as found from observations, is accordingly less than the theoretical value, since the ratio of the ranges at springs and neaps is reduced. This ratio is stated in the 'Admiralty Tide Tables for 1920' to be 1 : 2·73 on an average. If now $\theta$ be the phase of the lunar tide, let $(1-r)\theta$ be the phase of the solar tide, so that $r$ is 1/29. Let A be the amplitude of the lunar tidal current and $A\nu$ that of the solar tidal current. The total current is

$$A\{\cos\theta+\nu\cos(1-r)\theta\} = A(1+2\nu\cos r\theta+\nu^2)^{\frac{1}{2}}\cos\left(\theta-\tan^{-1}\frac{\sin r\theta}{1+\nu\cos r\theta}\right)$$

which is now expressed as a simple harmonic motion with a slowly varying amplitude and period. The amplitude at springs is $A(1+\nu)$. The dissipation is proportional to the cube of the current, and therefore to the cube of the amplitude. The ratio of the mean dissipation to the dissipation at springs is therefore the average of $(1+2\nu\cos r\theta+\nu^2)^{\frac{3}{2}}/(1+\nu)^3$. If $\nu^6$ be neglected, the numerator of this is $1+\frac{9}{4}\nu^2+\frac{9}{64}\nu^4$. Assuming that the ratio of the velocities is the same as that of the vertical ranges, we find that this fraction is equal to 0·51. Applying this correction to the spring tide dissipation, we find that the average dissipation is $1·1\times10^{19}$ ergs per second, 80 per cent. of what is required. It would give a secular acceleration of the moon of 7″ per century per century.

The agreement between the dissipation in shallow seas and that necessary to account for the lunar secular acceleration is much closer than the data would entitle us to expect. Two-thirds of that found takes place in the Bering Sea, the estimate for which may be incorrect by half its amount. What we are entitled to assert, however, is that this dissipation is certainly enough to account for a large fraction of the secular acceleration, and that there is nothing to prove that it is incapable of accounting for the whole of it.

It is uncertain whether the dissipation in any other coastal regions is notable in comparison with those already considered. The only partly enclosed areas not treated here that are of considerable size are some of those in the North-west Passage. There is an extensive shallow region off the coast of Patagonia, but it is in no way enclosed, being perfectly open to the Atlantic. Thus it is difficult to make any reliable inference about the currents in it. Many records of tidal currents along the coast are given, some reaching several knots, but all of them seem to refer to currents up rivers or near their mouths, where the general currents must be magnified, or to currents over bars and shoals; there seem to be no data about the currents more than a few miles out to sea.

The dissipation over local shallows like shoals and bars and in narrow bays and straits has been systematically ignored in this paper, except where it has been automatically taken into account in the determination of the excess of the entering over the issuing energy. The chief reason for this is the utter impossibility of finding it. The fjords on the west coasts of Norway, Greenland, and North and South America are innumerable, and in many of them, perhaps in all, there is a strong tidal current, so that the dissipation per unit area in these places must very much exceed that in any of the areas here treated. On the other hand, the total area must be less, and it is uncertain whether the increase in velocity is enough to counterbalance the decrease in area and make the total dissipation in these places comparable with that here found for the larger shallow seas. The same is true of shoals; though the agreement between the results given by the two methods of finding the dissipation in shoaly waters, as in the Yellow Sea and the Strait of Malacca, indicates that the shoals at any rate do not contribute to the dissipation an amount overwhelmingly greater than the normal places, for one method necessarily includes the effect of the shoals and the other systematically omits it. Along the open shore again there must be some dissipation; the currents there do not usually extend many miles out to sea, but they exist along a very long stretch of coast, and the aggregate dissipation in them may be appreciable.

The hypothesis that the secular acceleration of the moon is due to dissipation of energy in the tides in shallow coastal regions therefore seems capable of satisfying all the quantitative demands on it, and it is also free from objections that have been urged against other attempted explanations.* It therefore occupies a strong position.

## APPENDIX.

### *The Secular Change in the Obliquity of the Ecliptic.*

In consequence of the dissipation of energy in the diurnal tides there must be a couple always acting on the earth so as to tend to resist its angular motion about an axis in the plane of the orbit of the moon or the sun, as the case may be. If $\delta$ be the declination of the moon, the angular velocity of the earth about the diameter that points to the moon is $\Omega \sin \delta$, and the angular momentum about it is $C\Omega \sin \delta$, where C is the earth's moment of inertia. Let L be the couple about this diameter. Then the rate of dissipation of energy in the diurnal tide is $L\Omega \sin \delta$. Also the rate of change in the inclination is given by

$$\frac{d}{dt}(C\Omega \sin \delta) = L.$$

Now L must contain $\Omega \sin \delta$ as a factor, since it depends for its existence on the existence of the diurnal tide, whose coefficient is proportional to $\sin \delta$, and whose speed is equal to $\Omega-n$, which is sufficiently near to $\Omega$ for our present purpose. Hence the friction of the tide will have a damping effect on this component of the earth's rotation, which is not altered by the other couples acting, since none of these have a notable effect in diminishing the amplitude of the motion. If the average value of the dissipation is taken to be $5 \times 10^{18}$ ergs per second, this being rather more than was found in the South China Sea, and we remember that the average of $\sin^2 \delta$ is $\frac{1}{2} \sin^2 i$, where $i$ is the obliquity of the ecliptic, we find that the amplitude of $\Omega \sin \delta$ would be reduced to $1/e$ of its value in $2 \times 10^9$ years. This is of the same order as the probable age of the earth. If $\Omega$ remained constant this would show that the inclination of the Equator to the ecliptic would be reduced by $1''$ in about $2 \times 10^4$ years. Actually $\Omega$ is decreasing, so that if the rate were maintained it would be reduced to $1/e$ of its value in about $10^{10}$ years, a longer time than was found, on the assumption stated, to be enough for a similar reduction in the obliquity. Thus we can infer that the obliquity is at present diminishing, though there is no reason to believe that there has been any observable change in it in historic times. Even if there were as much dissipation in the diurnal tides as in the semi-diurnal ones this would hardly be possible.

---

* *Cf.* 'Monthly Notices of R.A.S.,' vol. lxxx., 1920, pp. 309–317.

[*Note added September 16.*—Mr. TAYLOR asks me to point out certain errata in his paper "Tidal Friction in the Irish Sea." On p. 2, $1-\frac{\gamma}{\sqrt{r}}$ is twice written for $1+\frac{\gamma}{\sqrt{r}}$; the correct form is used in equations (4) and (5). On p. 9, line 9, $T_1$ is Greenwich mean time of high water at full and change of the moon at the place considered, whereas the "establishment" is the *local* mean time of this event. In equation (16), $t+T_1$ should be $t-T_1$, and in equation (18), $t+T_0$ should be $t-T_0$. On pp. 19 and 20, $\sin^2 \phi_0$ is consistently written for $\sin 2\phi_0$; equation (33) is correct.

In this paper, as in TAYLOR'S, integrals over a period are always determined as if the current velocity and the tide height varied harmonically. This could be strictly correct only if the frictional force was proportional to the velocity, which is not the case. It appears, however, that the departure from the harmonic variation is not enough to produce any great alteration in these integrals.

I wish to express my thanks to Mr. H. W. BRABY for drawing the maps in this paper.]

# IX. *Plane Stress and Plane Strain in Bipolar Co-ordinates.*

*By* G. B. JEFFERY, *M.A., D.Sc., Fellow of University College, London.*

*Communicated by Prof.* L. N. G. FILON, *F.R.S.*

Received May 15,—Read June 24, 1920.

## § 1. INTRODUCTION.

THE problem of the equilibrium of an elastic solid under given applied forces is one of great difficulty and one which has attracted the attention of most of the great applied Mathematicians since the time of EULER. Unlike the kindred problems of hydrodynamics and electrostatics, it seems to be a branch of mathematical physics in which knowledge comes by the patient accumulation of special solutions rather than by the establishment of great general propositions. Nevertheless, the many and varied applications of this subject to practical affairs make it very desirable that these special solutions should be investigated, not only because of their intrinsic importance but also for the light which they often throw on the general problem. One of the most powerful methods of the mathematical physicist in the face of recalcitrant differential equations is to simplify his problem by reducing it to two dimensions. This simplification can only imperfectly be reproduced in the Nature of our three-dimensional world, but, in default of more general methods, it provides an invaluable weapon.

It was shown by AIRY* that in the two-dimensional case the stresses may be derived by partial differentiations from a single stress function, and it was shown later† that, in the absence of body forces, this stress function satisfies the linear partial differential equation of the fourth order $\nabla^4\chi = 0$, where $\nabla^4 = \nabla^2 . \nabla^2$, and $\nabla^2$ is the two-dimensional Laplacian $\partial^2/\partial x^2 + \partial^2/\partial y^2$.

It might have been expected that these results would have opened the way for a theory of two-dimensional elasticity of the same generality as the two-dimensional potential theory. This has not, however, been the case. This is due in part to the greater analytical difficulties which attend the discussion of the two-dimensional solutions of $\nabla^4\chi = 0$ as compared with $\nabla^2\chi = 0$. The analogues of many of the important properties of the simpler equation have yet to be discovered if they exist at all. Some progress has been made, and in this connection we may mention the work of J. H. MICHELL* who established a general theory of inversion which, with some important differences, follows the potential theory fairly closely.

---

* 'Brit. Assoc. Rep.,' 1862, p. 82.

† W. J. IBBETSON, 'Proc. Lond. Math. Soc.,' vol. xvii., 1886, p. 296. For a history of this part of the subject see LOVE'S 'Elasticity,' 2nd edition, p. 17.

No doubt the analytical difficulties have been the chief obstacle to progress, but perhaps the theory has not in recent years received attention which it would have received but for a certain physical difficulty. A truly two-dimensional elastic system is not so easy of realisation as might seem to be the case at first sight. If the stresses are everywhere parallel to the $xy$ plane and independent of $z$ there will in general be a varying displacement parallel to $z$. If the displacements are everywhere parallel to the $xy$ plane and independent of $z$ this can only be secured by the application of a stress $\widehat{zz}$ which varies from point to point and is perpendicular to the $xy$ plane. This difficulty was in a large measure removed by a theorem established by FILON, which has been called the theorem of generalised plane stress.† It states that if the average value of the stress $\widehat{zz}$ be taken throughout the thickness of a plate parallel to the $xy$ plane, then the ordinary two-dimensional theory will give accurately the *average* stresses through the thickness of the plate if the elastic constants of the material are modified. If $\lambda$, $\mu$ denote the true elastic constants, $\lambda$ must be replaced by $\lambda' = 2\lambda\mu/(\lambda+2\mu)$ while $\mu$ remains the same as before. This theorem attains an even greater importance when considered in the light of MICHELL'S theorem,‡ that if a plate bounded by any number of bounding curves is in equilibrium under forces in its plane applied over the boundaries, then, provided the forces applied over each boundary taken separately are in equilibrium, the stresses are everywhere independent of the elastic constants. The hypothesis that the average value of $\widehat{zz}$ vanishes throughout the plate, while certainly not accurately true in the majority of cases, will probably give a very close approximation in the case of a *thin* plate where parallel faces are unstressed.

In the light of this generalisation it is of considerable importance that the two-dimensional problem should be worked out more thoroughly. The two-dimensional solutions of $\nabla^4\chi = 0$ have been investigated in several systems of curvilinear co-ordinates. Owing to the special importance of the problem of the rectangular beam the solutions in Cartesian co-ordinates have naturally received a considerable amount of attention. MICHELL gave the general form of the stress-function in polar co-ordinates, thus opening the way for the solution of the problem of a plate bounded by two concentric circles, or an infinite plate containing a circular hole under any given tractions applied over its boundaries. In his lectures at University College, London, in 1912, Prof. FILON gave the complete solution of this problem determining the stresses and displacements when the stresses on the boundaries are expanded in Fourier series, and I am not aware that this solution has ever been published. An outline of the solution in elliptic co-ordinates is given in LOVE'S 'Elasticity.'*

---

* "The Inversion of Plane Stress," 'Proc. Lond. Math. Soc.,' 1901, vol. xxxiv., p. 134. Many of the results of the present paper can be obtained by an application of MICHELL'S methods, but it has proved more convenient to proceed on different lines.

† 'Roy. Soc. Phil. Trans.,' A, 1903, vol. 201, pp. 63–155.

‡ 'Proc. Lond. Math. Soc.,' vol. xxi., 1900, p. 100.

In this paper the complete solution is given for bipolar co-ordinates, for which the co-ordinate curves are co-axial circles. This solution enables us to treat the problems of an infinite plate containing two circular holes, a semi-infinite plate bounded by a straight edge and containing one circular hole, and a circular disc with an eccentric circular hole.

In the second Section the equations are expressed in bipolar co-ordinates and formulæ are established for the displacements in terms of the stress-function.

In the third Section the stress-function is obtained in a convenient form and the terms giving rise to many valued displacements are separated out.

The fourth Section is devoted to the determination of the coefficients in the stress-function when the tractions over the boundaries are given in Fourier series, and to an examination of the convergence of the resulting series. From the results established in this section it appears that the solution is complete, for the stress-function can always be uniquely determined when the tractions are given, provided that the applied forces taken as a whole are in equilibrium.

The remaining sections are occupied with the examination of some of the simpler applications of the theory. Section 5 gives the solution for a circular disc with an eccentric hole (or a cylinder with eccentric bore) when the two boundaries are under different hydrostatic pressures. It is found that the solution of this problem can be expressed in finite terms. An important particular case of this problem is discussed in Section 6, namely, a semi-infinite plate with a straight, unstressed boundary and a circular hole under a uniform normal pressure. This will give the stresses near a rivet hole while the hot plastic rivet is being forced home under pressure. This solution is interesting from another point of view, for if the ratio of the radius of the hole to its distance from the edge is suitably adjusted, the point of greatest tension will be on the straight edge while the point of greatest stress difference is on the circular boundary. It thus suggests a crucial test for the rival theories of rupture,—the greatest tension theory and the greatest stress-difference theory.

Section 7 deals with a semi-infinite plate with an unstressed circular hole under tension parallel to its straight edge. The solutions are in the form of infinite series, but the more important aspects of the problem are illustrated by numerical tables.

---

* 2nd edition, p. 259.

## § 2. THE CO-ORDINATES.

Let us take curvilinear co-ordinates defined by the conjugate functions

$$\alpha + i\beta = \log \frac{x + i(y + a)}{x + i(y - a)}, \quad \text{. . . . . . . . . . (1)}$$

where $x$, $y$ are Cartesian co-ordinates and $a$ is a positive real length. Solving for $x$, $y$, we have

$$x = \frac{a \sin \beta}{\cosh \alpha - \cos \beta}, \qquad y = \frac{a \sinh \alpha}{\cosh \alpha - \cos \beta}. \quad \text{. . . . . . . (2)}$$

Elements of arc measured along the normals to the curves $\alpha$, $\beta$ = *constant* are respectively $\delta\alpha/h$, $\delta\beta/h$, where

$$\frac{1}{h^2} = \left(\frac{\partial x}{\partial \alpha}\right)^2 + \left(\frac{\partial y}{\partial \alpha}\right)^2,$$

from which we have

$$h = (\cosh \alpha - \cos \beta)/a. \quad \text{. . . (3)}$$

The general scheme of co-ordinates is shown in fig. 1. If $O_1$, $O_2$ are the points $0, -a$ and $0, a$ respectively and P any point in the plane, and if the radii from $O_1$, $O_2$ to P are of lengths $r_1$, $r_2$ and are inclined at angles $\theta_1$, $\theta_2$ to the axis of $x$, then $\alpha = \log r_1/r_2$ and $\beta = \theta_1 - \theta_2$. The curves $\alpha$ = *constant* are a set of co-axial circles having $O_1$, $O_2$ for limiting points. The circles corresponding to positive values of $\alpha$ lie above the $x$-axis and those corresponding to negative values below, while the $x$-axis itself, which is the common radical axis, is given by $\alpha = 0$. The curves $\beta$ = *constant* are circles, or rather arcs of circles passing through $O_1$, $O_2$ and cutting the first set of circles orthogonally. On the right-hand side of the $y$-axis $\beta$ is positive and on the left-hand side negative, while on the $y$-axis $\beta = 0$, except on the segment $O_1O_2$, where $\beta = \pm\pi$. At infinity $\alpha = 0$, $\beta = 0$, and at $O_1$, $O_2$ we have $\alpha = -\infty$ and $+\infty$ respectively.

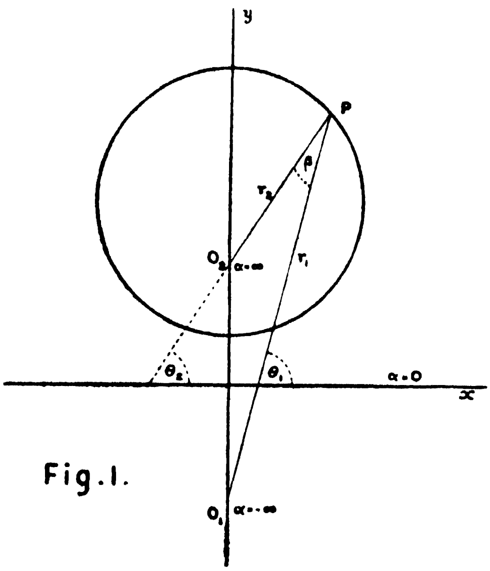


Fig. 1.

We have thus a set of co-ordinates adapted for the consideration of two-dimensional problems in which the region considered is—

(1) A finite region bounded internally by a circle and externally by a larger and non-concentric circle.

(2) A semi-infinite region bounded externally by a straight line and containing a circular hole.

(3) An infinite region containing two circular holes of any radii and centre distance.

If the displacements in the directions normal to the curves $\alpha$ and $\beta$ constant are $u$, $v$ respectively, the strains are given by[*]

$$e_{\alpha\alpha} = h\frac{\partial u}{\partial \alpha} - v\frac{\partial h}{\partial \beta}, \qquad e_{\beta\beta} = h\frac{\partial v}{\partial \beta} - u\frac{\partial h}{\partial \alpha},$$

$$e_{\alpha\beta} = \frac{\partial}{\partial \alpha}(hv) + \frac{\partial}{\partial \beta}(hu),$$

and the corresponding components of stress by

$$\left.\begin{aligned}\widehat{\alpha\alpha} &= \lambda\left(e_{\alpha\alpha}+e_{\beta\beta}\right)+2\mu e_{\alpha\alpha},\\ \widehat{\beta\beta} &= \lambda\left(e_{\alpha\alpha}+e_{\beta\beta}\right)+2\mu e_{\beta\beta},\\ \widehat{\alpha\beta} &= \mu e_{\alpha\beta}.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (4)$$

These stresses may be derived from a stress-function, so that in rectangular co-ordinates

$$\widehat{xx} = \frac{\partial^2\chi}{\partial y^2}, \qquad \widehat{xy} = -\frac{\partial^2\chi}{\partial x\,\partial y}, \qquad \widehat{yy} = \frac{\partial^2\chi}{\partial x^2}.$$

Transforming these equations to curvilinear co-ordinates we obtain

$$\left.\begin{aligned}\widehat{\alpha\alpha} &= h\frac{\partial}{\partial\beta}\left(h\frac{\partial\chi}{\partial\beta}\right)-h\frac{\partial h}{\partial\alpha}\frac{\partial\chi}{\partial\alpha},\\ \widehat{\beta\beta} &= h\frac{\partial}{\partial\alpha}\left(h\frac{\partial\chi}{\partial\alpha}\right)-h\frac{\partial h}{\partial\beta}\frac{\partial\chi}{\partial\beta},\\ \widehat{\alpha\beta} &= -h\frac{\partial^2(h\chi)}{\partial\alpha\,\partial\beta}+h\chi\frac{\partial^2 h}{\partial\alpha\,\partial\beta}.\end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (5)$$

We will usually find it convenient to deal with $h\chi$ instead of $\chi$ itself, and in our particular co-ordinates these equations become

$$\left.\begin{aligned}a\widehat{\alpha\alpha} &= \left\{(\cosh\alpha-\cos\beta)\frac{\partial^2}{\partial\beta^2}-\sinh\alpha\frac{\partial}{\partial\alpha}-\sin\beta\frac{\partial}{\partial\beta}+\cosh\alpha\right\}(h\chi).\\ a\widehat{\beta\beta} &= \left\{(\cosh\alpha-\cos\beta)\frac{\partial^2}{\partial\alpha^2}-\sinh\alpha\frac{\partial}{\partial\alpha}-\sin\beta\frac{\partial}{\partial\beta}+\cosh\beta\right\}(h\chi).\\ a\widehat{\alpha\beta} &= -(\cosh\alpha-\cos\beta)\frac{\partial^2(h\chi)}{\partial\alpha\,\partial\beta}.\end{aligned}\right\} \quad . \quad . \quad . \quad (6)$$

We may note that

$$a\left(\widehat{\alpha\alpha}-\widehat{\beta\beta}\right) = (\cosh\alpha-\cos\beta)\left(\frac{\partial^2}{\partial\beta^2}-\frac{\partial^2}{\partial\alpha^2}+1\right)(h\chi), \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (7)$$

---

[*] LOVE, 'Theory of Elasticity,' p. 54.

so that if $h\chi$ and its second differential coefficients are finite at infinity ($\alpha = 0$. $\beta = 0$) we have there $a\widehat{\alpha\alpha} = a\widehat{\beta\beta} = h\chi$ and $\widehat{\alpha\beta} = 0$.

In the absence of body forces the stress-function satisfies $\nabla^4\chi = 0$. In curvilinear co-ordinates we have

$$\nabla^2 \equiv h^2\left(\frac{\partial^2}{\partial\alpha^2} + \frac{\partial^2}{\partial\beta^2}\right),$$

and, taking $h\chi$ as the dependent variable, we have in our co-ordinates

$$a\nabla^2\chi = \left\{(\cosh\alpha - \cos\beta)\left(\frac{\partial^2}{\partial\alpha^2} + \frac{\partial^2}{\partial\beta^2}\right) - 2\sinh\alpha\frac{\partial}{\partial\alpha} - 2\sin\beta\frac{\partial}{\partial\beta} + \cosh\alpha + \cos\beta\right\}(h\chi).$$

Repeating the operator, a little reduction leads us to the following transformation for $\nabla^4\chi = 0$:

$$\left(\frac{\partial^4}{\partial\alpha^4} + 2\frac{\partial^4}{\partial\alpha^2\partial\beta^2} + \frac{\partial^4}{\partial\beta^4} - 2\frac{\partial^2}{\partial\alpha^2} + 2\frac{\partial^2}{\partial\beta^2} + 1\right)(h\chi) = 0. \quad . \quad . \quad . \quad . \quad . \quad (8)$$

Thus by considering $h\chi$ instead of $\chi$ we have a linear equation with constant coefficients.

Before proceeding to the discussion of its solutions, we must investigate the method of determining the displacements corresponding to a given stress-function, in order that we may ascertain whether and under what conditions these are single-valued. This is particularly necessary in our case, as one of the co-ordinates, $\beta$, is itself many-valued.

Adding and subtracting the first two equations (4), and leaving the third as it stands, substituting for the stresses in terms of the stress-function, and for the strains in terms of the displacements, we obtain the following three equations:—

$$\frac{\partial}{\partial\alpha}\left\{\frac{\partial\chi}{\partial\alpha} - 2(\lambda+\mu)\frac{u}{h}\right\} + \frac{\partial}{\partial\beta}\left\{\frac{\partial\chi}{\partial\beta} - 2(\lambda+\mu)\frac{v}{h}\right\} = 0, \quad . \quad . \quad . \quad . \quad . \quad (9)$$

$$\frac{\partial}{\partial\alpha}\left\{h^2\frac{\partial\chi}{\partial\alpha} + 2\mu hu\right\} - \frac{\partial}{\partial\beta}\left\{h^2\frac{\partial\chi}{\partial\beta} + 2\mu hv\right\} = 0, \quad . \quad . \quad . \quad . \quad . \quad . \quad (10)$$

$$\frac{\partial}{\partial\alpha}\left\{h^2\frac{\partial\chi}{\partial\beta} + 2\mu hv\right\} + \frac{\partial}{\partial\beta}\left\{h^2\frac{\partial\chi}{\partial\alpha} + 2\mu hu\right\} = 0. \quad . \quad . \quad . \quad . \quad . \quad . \quad (11)$$

From the last two of these it appears that we may define a new function P such that

$$\frac{\partial P}{\partial\alpha} = h^2\frac{\partial\chi}{\partial\beta} + 2\mu hv, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (12)$$

$$\frac{\partial P}{\partial\beta} = h^2\frac{\partial\chi}{\partial\alpha} + 2\mu hu, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (13)$$

$$\nabla^2 P = 0,$$

and we have still to satisfy (9). Substituting for $u$, $v$ in terms of P we have

$$h^2\frac{\partial}{\partial\alpha}\left(\frac{1}{h^2}\frac{\partial P}{\partial\beta}\right)+h^2\frac{\partial}{\partial\beta}\left(\frac{1}{h^2}\frac{\partial P}{\partial\alpha}\right)=\frac{\lambda+2\mu}{\lambda+\mu}\nabla^2\chi,$$

which may be re-arranged thus—

$$\frac{\partial}{\partial\alpha}\left\{\frac{\lambda+2\mu}{\lambda+\mu}\frac{\partial\chi}{\partial\alpha}-\frac{1}{h^2}\frac{\partial P}{\partial\beta}\right\}+\frac{\partial}{\partial\beta}\left\{\frac{\lambda+2\mu}{\lambda+\mu}\frac{\partial\chi}{\partial\beta}-\frac{1}{h^2}\frac{\partial P}{\partial\alpha}\right\}=0.$$

It follows that a function Q exists such that

$$h^2\frac{\partial Q}{\partial\alpha}=\frac{\partial P}{\partial\alpha}-\frac{\lambda+2\mu}{\lambda+\mu}h^2\frac{\partial\chi}{\partial\beta}\quad\ldots\ldots\ldots\ldots\quad(14)$$

$$h^2\frac{\partial Q}{\partial\beta}=-\frac{\partial P}{\partial\beta}+\frac{\lambda+2\mu}{\lambda+\mu}h^2\frac{\partial\chi}{\partial\alpha}\quad\ldots\ldots\ldots\ldots\quad(15)$$

Eliminating P by differentiating with regard to $\beta$ and $\alpha$ respectively, and adding, we have

$$\frac{\partial^2(hQ)}{\partial\alpha\,\partial\beta}-Q\frac{\partial^2 h}{\partial\alpha\,\partial\beta}=\frac{\lambda+2\mu}{2(\lambda+\mu)}\left\{\frac{1}{h}\frac{\partial}{\partial\alpha}\left(h^2\frac{\partial\chi}{\partial\alpha}\right)-\frac{1}{h}\frac{\partial}{\partial\beta}\left(h^2\frac{\partial\chi}{\partial\beta}\right)\right\},$$

which becomes in our co-ordinates

$$\frac{\partial^2(hQ)}{\partial\alpha\,\partial\beta}=\frac{\lambda+2\mu}{2(\lambda+\mu)}\left\{\frac{\partial^2(h\chi)}{\partial\alpha^2}-\frac{\partial^2(h\chi)}{\partial\beta^2}-h\chi\right\}\quad\ldots\ldots\ldots\quad(16)$$

There is, however, a further condition to be satisfied by Q corresponding to the condition $\nabla^2 P=0$. Differentiating (14) and (15) with regard to $\alpha$, $\beta$ respectively, and subtracting we have

$$\frac{\partial}{\partial\alpha}\left(h^2\frac{\partial Q}{\partial\alpha}\right)-\frac{\partial}{\partial\beta}\left(h^2\frac{\partial Q}{\partial\beta}\right)=-\frac{\lambda+2\mu}{\lambda+\mu}\left\{\frac{\partial}{\partial\alpha}\left(h^2\frac{\partial\chi}{\partial\beta}\right)+\frac{\partial}{\partial\beta}\left(h^2\frac{\partial\chi}{\partial\alpha}\right)\right\}.$$

or in our co-ordinates

$$\frac{\partial^2}{\partial\alpha^2}(hQ)-\frac{\partial^2}{\partial\beta^2}(hQ)-hQ=-\frac{2(\lambda+2\mu)}{\lambda+\mu}\frac{\partial^2(h\chi)}{\partial\alpha\,\partial\beta}\quad\ldots\ldots\ldots\quad(17)$$

These two equations connecting Q and $\chi$ are consistent, for, if we eliminate Q by appropriate differential operators, we have

$$\left\{\frac{\partial^2}{\partial\alpha^2}-\frac{\partial^2}{\partial\beta^2}-1\right\}^2(h\chi)=-4\frac{\partial^4(h\chi)}{\partial\alpha^2\,\partial\beta^2},$$

which is readily seen to be identical with the condition $\nabla^4\chi=0$, as given in (8). It is obvious that $hQ$ satisfies the same differential equation, and hence it also is a solution of $\nabla^4 Q=0$.

We have therefore from (16)

$$hQ = \frac{\lambda+2\mu}{2(\lambda+\mu)}\iint\left\{\frac{\partial^2(h\chi)}{\partial\alpha^2}-\frac{\partial^2(h\chi)}{\partial\beta^2}-h\chi\right\}d\alpha\,d\beta, \quad . \quad . \quad . \quad . \quad . \quad (18)$$

and from (12), (13) and (14), (15)

$$2\mu u = \frac{\mu}{\lambda+\mu}h\frac{\partial\chi}{\partial\alpha}-h\frac{\partial Q}{\partial\beta}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (19)$$

$$2\mu v = \frac{\mu}{\lambda+\mu}h\frac{\partial\chi}{\partial\beta}+h\frac{\partial Q}{\partial\alpha}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (20)$$

It is readily seen that these equations determine $u$ and $v$ apart possibly from rigid body displacements, for, although owing to the double integration an arbitrary function of $\alpha$ and an arbitrary function of $\beta$ will appear in $hQ$, these will be determined by (17), except for functions of $\alpha$ or $\beta$, which make its left-hand side vanish identically. The only possible arbitrary terms in $hQ$ are therefore given by $hQ = a\text{A}(\cosh\alpha+\cos\beta)+\text{B}(\cosh\alpha-\cos\beta)+\text{C}a\sinh\alpha+\text{D}a\sin\beta$, or

$$Q = \text{A}r^2+a\text{B}+\text{C}y+\text{D}x$$

where $r$ is the distance from the origin. These give rise to terms in $u$, $v$ corresponding to motions of pure translation and rigid body rotation about the origin.

## § 3. THE STRESS-FUNCTION.

Turning now to the consideration of the possible forms for the stress-function in these co-ordinates, we note that the differential equation (8) can readily be solved by the ordinary method, and that its general solution is

$$h\chi = e^{\alpha}\phi_1(\alpha+i\beta)+e^{-\alpha}\phi_2(\alpha+i\beta)+e^{\alpha}\phi_3(\alpha-i\beta)+e^{-\alpha}\phi_4(\alpha-i\beta).$$

If we seek a solution of the type $h\chi = f(\alpha)\cos n\beta$ or $f(\alpha)\sin n\beta$, (8) shows that the differential equation for $f(\alpha)$ is

$$\left(\frac{d^4}{d\alpha^4}-2(n^2+1)\frac{d^2}{d\alpha^2}+n^4-2n^2+1\right)f(\alpha) = 0,$$

the solution of which is

$$f(\alpha) = \text{A}_n\cosh(n+1)\alpha+\text{B}_n\cosh(n-1)\alpha+\text{C}_n\sinh(n+1)\alpha+\text{D}_n\sinh(n-1)\alpha,$$

unless $n = 0$ or $1$. In the latter case we have

$$f(\alpha) = \text{A}_1\cosh 2\alpha+\text{B}_1+\text{C}_1\sinh 2\alpha+\text{D}_1\alpha,$$

and when $n = 0$

$$f(\alpha) = \text{A}_0\cosh\alpha+\text{B}_0\alpha\cosh\alpha+\text{C}_0\sinh\alpha+\text{D}_0\alpha\sinh\alpha.$$

If we now seek solutions for which $h\chi$ is a multiple of $\sinh n\alpha$ or $\cosh n\alpha$, we find the following solutions which are not included above—

$$h\chi = (\mathrm{E}\cos\beta+\mathrm{F}\sin\beta+\mathrm{G}\cosh\alpha+\mathrm{H}\sinh\alpha)\,\beta.$$

Since any constant multiples of $x$, $y$ and any constant may be added to $\chi$ without affecting the stresses, it follows from (2) that any multiples of

$$\sinh\alpha,\ \sin\beta \text{ or } \cosh\alpha-\cos\beta$$

may be added to $h\chi$. This allows us to take the coefficients of $\cosh\alpha$, $\sinh\alpha$ and $\sin\beta$ as zero. We have then the following general expression for $h\chi$:—

$$\begin{aligned} h\chi = {} & (\mathrm{E}\cos\beta+\mathrm{F}\sin\beta+\mathrm{G}\cosh\alpha+\mathrm{H}\sinh\alpha)\,\beta \\ & +(\mathrm{B}_0\cosh\alpha+\mathrm{D}_0\sinh\alpha)\,\alpha \\ & +(\mathrm{A}_1\cosh 2\alpha+\mathrm{B}_1+\mathrm{C}_1\sinh 2\alpha+\mathrm{D}_1\alpha)\cos\beta \\ & +(\mathrm{A}'_1\cosh 2\alpha+\mathrm{C}'_1\sinh 2\alpha+\mathrm{D}'_1\alpha)\sin\beta \\ & +\sum_{n=2}^{\infty}\left\{\begin{array}{r} [\mathrm{A}_n\cosh(n+1)\alpha+\mathrm{B}_n\cosh(n-1)\alpha+\mathrm{C}_n\sinh(n+1)\alpha \\ +\mathrm{D}_n\sinh(n-1)\alpha]\cos n\beta \\ +[\mathrm{A}'\cosh(n+1)\alpha+\mathrm{B}'_n\cosh(n-1)\alpha+\mathrm{C}'_n\sinh(n+1)\alpha \\ +\mathrm{D}'_n\sinh(n-1)\alpha]\sin n\beta. \end{array}\right\} \end{aligned} \quad . \; . \; . \quad (21)$$

We have now to determine whether the displacements corresponding to this stress-function are single valued or not. The function $(h\mathrm{Q})$ is easily obtained by simple integration from (16), and the arbitrary functions thus appearing can be determined by the aid of (17). We have

$$\begin{aligned} -\frac{\lambda+\mu}{\lambda+2\mu}(h\mathrm{Q}) = {} & (\mathrm{E}\cos\beta+\mathrm{F}\sin\beta+\mathrm{G}\cosh\alpha+\mathrm{H}\sinh\alpha)\,\alpha \\ & -(\mathrm{B}_0\cosh\alpha+\mathrm{D}_0\sinh\alpha)\,\beta \\ & -(\mathrm{A}_1\sinh 2\alpha+\mathrm{C}_1\cosh 2\alpha+\mathrm{D}'_1\beta)\sin\beta \\ & +(\mathrm{A}'_1\sinh 2\alpha+\mathrm{C}'_1\cosh 2\alpha-\mathrm{D}_1\beta)\cos\beta \\ & +\sum_{n=2}^{\infty}\left\{\begin{array}{r} [\mathrm{A}'_n\sinh(n+1)\alpha+\mathrm{B}'_n\sinh(n-1)\alpha+\mathrm{C}'_n\cosh(n+1)\alpha \\ +\mathrm{D}'_n\cosh(n-1)\alpha]\cos n\beta \\ -[\mathrm{A}_n\sinh(n+1)\alpha+\mathrm{B}_n\sinh(n-1)\alpha+\mathrm{C}_n\cosh(n+1)\alpha \\ +\mathrm{D}_n\cosh(n-1)\alpha]\sin n\beta. \end{array}\right\} \end{aligned} \quad . \; . \; . \quad (22)$$

It is clear, from the general expressions for $h\chi$ and $h\mathrm{Q}$, that the only terms which can possibly give rise to many-valued displacements are

$$\begin{aligned} h\chi = {} & (\mathrm{E}\cos\beta+\mathrm{F}\sin\beta+\mathrm{G}\cosh\alpha+\mathrm{H}\sinh\alpha)\,\beta \\ & +(\mathrm{B}_0\cosh\alpha+\mathrm{D}_0\sinh\alpha+\mathrm{D}_1\cos\beta+\mathrm{D}'_1\sin\beta)\,\alpha, \end{aligned}$$

and the corresponding terms in $hQ$

$$-\frac{\lambda+\mu}{\lambda+2\mu}hQ = (E\cos\beta+F\sin\beta+G\cosh\alpha+H\sinh\alpha)\,\alpha$$
$$-(B_0\cosh\alpha+D_0\sinh\alpha+D_1\cos\beta+D'_1\sin\beta)\,\beta.$$

From (19) and (20) we may now find the corresponding displacements $u$, $v$. Each of these is found to contain a multiple of the many-valued co-ordinate $\beta$. Equating the coefficients of these terms to zero we have the following relations:—

$$\left.\begin{array}{rr} E+G=0, & B_0+D_1=0, \\ \mu F-(\lambda+2\mu)\,D_0=0, & \mu H+(\lambda+2\mu)\,D'_1=0. \end{array}\right\} \quad . \quad . \quad . \quad . \quad . \quad (23)$$

We shall now show that these early terms correspond to the resultant of the forces and couple applied over the boundaries. For this purpose we shall require the following elementary forms of the stress-function :—

(1) For an isolated force X applied at the origin in the direction of the $x$-axis

$$\chi = -(2\pi)^{-1}X\,(y\theta-\nu x\log r)$$

where $r$, $\theta$ as usual denote polar co-ordinates and $\nu = \mu/(\lambda+2\mu)$.

(2) For an isolated force Y applied at the origin in the direction of the $y$-axis

$$\chi = (2\pi)^{-1}Y\,(x\theta+\nu y\log r).$$

(3) For a point couple of moment L applied at the origin in a positive sense

$$\chi = -(2\pi)^{-1}L\theta.$$

(4) For a centre of pressure radiating uniformly from the origin

$$\chi = \log r,$$

Inserting the relations (23) necessary to ensure single-valued displacements our early terms become

$$h\chi = G(\cosh\alpha-\cos\beta)\,\beta+\beta_0(\cosh\alpha-\cos\beta)\,\alpha$$
$$+F\,(\beta\sin\beta+\nu\alpha\sinh\alpha)+H\,(\beta\sinh\alpha-\nu\alpha\sin\beta)$$

or

$$\chi = aG\beta+aB_0\alpha+F\,(x\beta+\nu y\alpha)+H\,(y\beta-\nu x\alpha). \quad . \quad . \quad . \quad . \quad . \quad (24)$$

Now $aG\beta = aG\,(\theta_1-\theta_2)$ and hence this term represents a couple of moment $2\pi aG$ applied at $\alpha = \infty$ and an equal and opposite couple applied at $\alpha = -\infty$. The term $a\beta_0\alpha$ represents two equal and opposite centres of radial pressure at these same points.

We also have

$$H(y\beta - \nu x\alpha) = H\{(y+a)\theta_1 - \nu x \log r_1\} - H\{(y-a)\theta_2 - \nu x \log r_2\}$$
$$-aH(\theta_1+\theta_2).$$

This corresponds to a force $2\pi H$ applied at $\alpha = +\infty$ parallel to the $x$-axis and an equal and opposite force applied at $\alpha = -\infty$ (thus forming a couple of moment $4\pi a H$) and point couples each of moment $2\pi a H$ applied at these same points (see fig. 2).

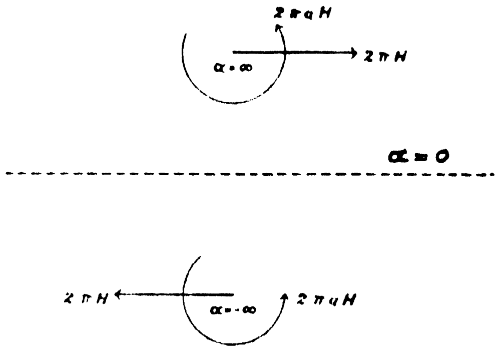


Fig. 2.

Finally

$$F(x\beta + \nu y\alpha) = F\{x\theta_1 + \nu(y+a)\log r_1\}$$
$$-F\{x\theta_2 + \nu(y-a)\log r_2\}$$
$$-a\nu F \log r_1 r_2.$$

This corresponds to forces each equal to $2\pi F$, acting at the points $\alpha = \pm\infty$ and each directed towards the origin, together with two equal like centres of uniform pressure at the same points. This brings to light a new solution corresponding to the last term.

Expressed in our co-ordinates we have

$$\log r_1 r_2 = 2\log(2a) - 2\log(\cosh\alpha - \cos\beta),$$

and the corresponding form of $h\chi$ is, apart from constants,

$$h\chi = (\cosh\alpha - \cos\beta)\log(\cosh\alpha - \cos\beta).$$

It is easily seen that this can be expanded in a Fourier series which is included in our general expression for $h\chi$, but that the expansion is different on opposite sides of the line $\alpha = 0$. For this reason we shall find it convenient to include a term of this form whenever the region under consideration includes parts above and below the axis of $x$, *i.e.*, when it is bounded by two circles neither of which encloses the other.

It will be noted that, taken together, the early terms allow for the most general resultant forces acting over the two circular boundaries enclosing the two points $\alpha = +\infty$, $\alpha = -\infty$, subject to the condition that the forces acting over the two boundaries considered together form a system in equilibrium. If it is desired to investigate problems for which this condition is not satisfied we can readily obtain the necessary additional solutions. They will be

$$\left.\begin{aligned} \chi &= (y+a)\theta_1 - \nu x \log r_1 \\ \chi &= x\theta_1 + \nu(y+\alpha)\log r_1 \\ \chi &= \theta_1 \end{aligned}\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (25)$$

corresponding to forces and couple applied at $\alpha = -\infty$, and similar terms in $\theta_2$, $\log r_2$ corresponding to forces and couple applied at $\alpha = +\infty$. The corresponding forms of $h\chi$ can be expanded in series which are included in our general form, but here again the expansions are different on opposite sides of $\alpha = 0$ and diverge for $\alpha = 0$, $\beta = 0$ together, *i.e.*, at infinity. This divergence corresponds to the obvious fact that forces or couples must be applied at infinity to maintain equilibrium.

Owing to difficulties of this kind we shall find it convenient to insert the appropriate terms corresponding to the resultant force and couple over a boundary and to investigate the stress-function corresponding to the remaining applied forces which will be in statical equilibrium for each boundary.

Let us write for brevity

$$\left.\begin{aligned}\phi_n(\alpha) &= \mathrm{A}_n \cosh(n+1)\alpha + \mathrm{B}_n \cosh(n-1)\alpha + \mathrm{C}_n \sinh(n+1)\alpha + \mathrm{D}_n \sinh(n-1)\alpha \\ \psi_n(\alpha) &= \mathrm{A}'_n \cosh(n+1)\alpha + \mathrm{B}'_n \cosh(n-1)\alpha + \mathrm{C}'_n \sinh(n+1)\alpha + \mathrm{D}_n{}' \sinh(n-1)\alpha,\end{aligned}\right\} \qquad (26)$$

if $n \geqq 2$ and

$$\left.\begin{aligned}\phi_1(\alpha) &= \mathrm{A}_1 \cosh 2\alpha + \mathrm{B}_1 + \mathrm{C}_1 \sinh 2\alpha \\ \psi_1(\alpha) &= \mathrm{A}'_1 \cosh 2\alpha + \mathrm{C}'_1 \sinh 2\alpha.\end{aligned}\right\} \quad . \ . \ . \ . \ . \ . \ . \quad (27)$$

Setting aside the terms corresponding to the resultant forces and couples over the separate boundaries we have

$$\begin{aligned}h\chi = {} & \{\mathrm{B}_0\alpha + \mathrm{K}\log(\cosh\alpha - \cos\beta)\}(\cosh\alpha - \cos\beta) \\ & + \sum_{n=1}^{\infty}\{\phi_n(\alpha)\cos n\beta + \psi_n(\alpha)\sin n\beta\} \quad . \ . \ . \ . \ . \ . \ . \ . \quad (28)\end{aligned}$$

where the term in K may be omitted when the region considered lies entirely on one side of the line

## § 4. BOUNDARY CONDITIONS.

Let us consider a plate bounded by two curves $\alpha = \alpha_1, \alpha_2$. We may suppose $\alpha_1 > \alpha_2$ and $\alpha_1 > 0$. Then, if $\alpha_2 > 0$ we have a finite plate bounded internally and externally by circles which are not concentric, if $\alpha_2 < 0$ we have an infinite plate containing two circular holes, and by suitably choosing the values of $a$, $\alpha_1$, $\alpha_2$ we can make the circular boundaries in either case of any desired radius and centre distance. In particular if $\alpha_2 = 0$, we have a semi-infinite plate bounded by a straight edge and containing a circular hole. Suppose that such a plate is in equilibrium under given normal and tangential forces applied over the boundaries $\alpha = \alpha_1, \alpha_2$, so that we are given over $\alpha = \alpha_1$,

$$\left.\begin{aligned}\widehat{a\alpha\beta} &= a_0 + \sum_{1}^{\infty}(a_n \cos n\beta + b_n \sin n\beta), \\ \widehat{a\alpha\alpha} &= c_0 + \sum_{1}^{\infty}(c_n \cos n\beta + d_n \sin n\beta),\end{aligned}\right\} \quad . \ . \ . \ . \ . \ . \ . \quad (29)$$

while over $\alpha = \alpha_2$ we have similar expansions in which $a_0$, $a_n$, $b_n$, $c_0$, $c_n$, $d_n$ are replaced by $a'_0$, $a'_n$, $b'_n$, $c'_0$, $c'_n$, $d'_n$.

If the tractions applied over the circle $\alpha = \alpha_1$ are statically equivalent to forces X, Y at its centre, and a couple of moment L, then

$$X = \int_0^{2\pi} \left\{ \widehat{\alpha\alpha} \frac{\partial x}{\partial \alpha} - \widehat{\alpha\beta} \frac{\partial y}{\partial \alpha} \right\} d\beta,$$

$$Y = \int_0^{2\pi} \left\{ \widehat{\alpha\alpha} \frac{\partial y}{\partial \alpha} + \widehat{\alpha\beta} \frac{\partial x}{\partial \alpha} \right\} d\beta,$$

$$L = -\frac{a^2}{\sinh \alpha_1} \int_0^{2\pi} \frac{\widehat{\alpha\beta}\, d\beta}{\cosh \alpha - \cos \beta}.$$

The coefficients of $\widehat{\alpha\alpha}$, $\widehat{\beta\beta}$ can readily be expanded in Fourier series. We have, in fact, since $\alpha_1 > 0$,

$$\frac{\partial x}{\partial \alpha} = -a \frac{\sinh \alpha \sin \beta}{(\cosh \alpha - \cos \beta)^2} = -2a \sum_1^\infty n e^{-n\alpha_1} \sin n\beta,$$

$$\frac{\partial y}{\partial \alpha} = -a \frac{(\cosh \alpha \cos \beta - 1)}{(\cosh \alpha - \cos \beta)^2} = -2a \sum_1^\infty n e^{-n\alpha_1} \cos n\beta,$$

and

$$\sinh \alpha_1 (\cosh \alpha_1 - \cos \beta)^{-1} = 1 + 2 \sum_1^\infty e^{-n\alpha_1} \cos n\beta.$$

Substituting these and the expansions for $\widehat{\alpha\alpha}$, $\widehat{\alpha\beta}$ in the expressions for X, Y, L, and integrating, we have

$$X = 2\pi \sum_1^\infty n (a_n - d_n) e^{-n\alpha_1},$$

$$Y = -2\pi \sum_1^\infty n (b_n + c_n) e^{-n\alpha_1},$$

$$L = -2\pi a \operatorname{cosech}^2 \alpha_1 \sum_1^\infty a_n e^{-n\alpha_1}.$$

The corresponding components of the resultant of the forces applied over $\alpha = \alpha_2$ can be obtained in a similar way. We must, however, remember that in this case the forces act from that side of the boundary for which $\alpha < \alpha_2$, whereas in the case of the first boundary they acted from the side for which $\alpha < \alpha_1$. We obtain, if $\alpha_2 > 0$,

$$X' = -2\pi \sum_1^\infty n (a'_n - d'_n) e^{-n\alpha_2},$$

$$Y' = 2\pi \sum_1^\infty n (b'_n + c'_n) e^{-n\alpha_2},$$

$$L' = 2\pi a \operatorname{cosech}^2 \alpha_2 \sum_1^\infty a'_n e^{-n\alpha_2}.$$

If $\alpha_2 < 0$ there are some differences of sign owing to the different Fourier expansions for the direction cosines. We have

$$X' = -2\pi \sum_1^\infty n\,(a'_n + d'_n)\, e^{n\alpha_2},$$

$$Y' = -2\pi \sum_1^\infty n\,(b'_n - c'_n)\, e^{n\alpha_2},$$

$$L' = -2\pi a \operatorname{cosech}^2\alpha_2 \sum_1^\infty a'_n e^{n\alpha_2}.$$

Hence, if the forces acting on each boundary are statically in equilibrium, we have

$$\left.\begin{array}{ll} \sum_1^\infty n\,(a_n - d_n)\, e^{-n\alpha_1} = 0, & \sum_1^\infty n\,(b_n + c_n)\, e^{-n\alpha_1} = 0, \\ \sum_1^\infty a_n e^{-n\alpha_1} = 0, & \sum_1^\infty a'_n e^{\pm n\alpha_2} = 0. \end{array}\right\} \quad . \quad . \quad . \quad . \quad (30)$$

with, if $\alpha_2 > 0$,

$$\sum_1^\infty n\,(a'_n - d'_n)\, e^{-n\alpha_2} = 0, \qquad \sum_1^\infty n\,(b'_n + c'_n)\, e^{-n\alpha_2} = 0, \quad . \quad . \quad . \quad . \quad (31)$$

or, if $\alpha_2 < 0$,

$$\sum_1^\infty n\,(a'_n + d'_n)\, e^{n\alpha_2} = 0, \qquad \sum_1^\infty n\,(b'_v - c'_n)\, e^{n\alpha_2} = 0. \quad . \quad . \quad . \quad . \quad . \quad (32)$$

We will now show that it is possible to determine a stress-function of the form (28) which gives the appropriate stresses over $\alpha = \alpha_1, \alpha_2$, and which gives no stress at infinity if the region considered extends so far.

By the aid of (6) we can calculate the stresses corresponding to the stress-function (28). We obtain

$$\begin{aligned} 2a\widehat{\alpha\alpha} = {} & K\,(1 - 2\cosh^2\alpha) - 2B_0 \sinh\alpha\cosh\alpha + 2\phi_1(\alpha) \\ & + 2\,(K\cosh\alpha + B_0\sinh\alpha)\cos\beta - K\cos 2\beta \\ & + \sum_{n=1}^\infty \left\{ \begin{array}{l} [(n+1)(n+2)\,\phi_{n+1}(\alpha) - 2\cosh\alpha\,(n^2-1)\,\phi_n(\alpha) \\ \qquad\qquad + (n-1)(n-2)\,\phi_{n-1}(\alpha)]\cos n\beta \\ + [(n+1)(n+2)\,\psi_{n+1}(\alpha) - 2\cosh\alpha\,(n^2-1)\,\psi_n(\alpha) \\ \qquad\qquad + (n-1)(n-2)\,\psi_{n-1}(\alpha)]\sin n\beta \\ -2\sinh\alpha\,[\phi'_n(\alpha)\cos n\beta + \psi'_n(\alpha)\sin n\beta]. \end{array} \right\} \end{aligned}$$

and

$$\begin{aligned} 2a\widehat{\alpha\beta} = {} & \psi'_1(\alpha) - 2\,(K\sinh\alpha + B_0\cosh\alpha)\sin\beta + B_0\sin 2\beta \\ & + \sum_{n=1}^\infty \left\{ \begin{array}{l} [(n+1)\,\psi'_{n+1}(\alpha) - 2\cosh\alpha n\psi'_n(\alpha) + (n-1)\,\psi'_{n-1}(\alpha)]\cos n\beta, \\ -[(n+1)\,\phi'_{n+1}(\alpha) - 2\cosh\alpha n\phi'_n(\alpha) + (n-1)\,\phi'_{n-1}(\alpha)]\sin n\beta. \end{array} \right\} \end{aligned}$$

Identifying these with (29), we have the following relations from which to obtain the coefficients:—

$$\left.\begin{array}{l}
\phi_1(\alpha_1) = 2c_0+2\mathrm{B}_0 \sinh\alpha_1 \cos\alpha_1+2\mathrm{K}\cosh^2\alpha_1-\mathrm{K} \\
1.2.3.\phi_2(\alpha_1) = 2c_1-2(\mathrm{K}\cosh\alpha_1+\mathrm{B}_0\sinh\alpha_1)+2\sinh\alpha_1\,\phi'_1(\alpha_1) \\
2.3.4.\phi_3(\alpha_1)-2\cosh\alpha_1\,.\,1.2.3.\phi_2(\alpha_1) = 4c_2+2\mathrm{K}+4\sinh\alpha_1\,\phi'_2(\alpha_1) \\
\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot \\
n(n+1)(n+2)\phi_{n+1}(\alpha_1)-2\cosh\alpha_1(n-1)(n)(n+1)\phi_n(\alpha_1)+(n-2)(n-1)(n)\phi_{n-1}(\alpha_1) \\
\qquad\qquad = 2nc_n+2n\sinh\alpha_1\,\phi'_n(\alpha_1) \qquad\qquad (n \geqq 3)
\end{array}\right\} \quad (33)$$

$$\left.\begin{array}{l}
n(n+1)(n+2)\psi_{n+1}(\alpha_1)-2\cosh\alpha_1(n-1)(n)(n+1)\psi_n(\alpha_1)+(n-2)(n-1)(n)\psi_{n-1}(\alpha_1) \\
\qquad\qquad = 2nd_n+2n\sinh\alpha_1\,\psi'_n(\alpha_1) \qquad\qquad (n \geqq 1)
\end{array}\right\} \quad .\ .\ .\ (34)$$

$$\left.\begin{array}{l}
\psi'_1(\alpha_1) = 2a_0 \\
(n+1)\psi'_{n+1}(\alpha_1)-2\cosh\alpha_1\,n\psi'_n(\alpha_1)+(n-1)\psi'_{n-1}(\alpha_1) = 2a_n \qquad (n \geqq 1)
\end{array}\right\} \quad .\ (35)$$

$$\left.\begin{array}{l}
2\phi'_2(\alpha_1)-2\cosh\alpha_1\,\phi'_1(\alpha_1) = -2b_1-2(\mathrm{K}\sinh\alpha_1+\mathrm{B}_0\cosh\alpha_1) \\
3\phi'_3(\alpha_1)-4\cosh\alpha_1\,\phi'_2(\alpha_1)+\phi'_1(\alpha_1) = -2b_2+\mathrm{B}_0 \\
\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot\quad\cdot \\
(n+1)\phi'_{n+1}(\alpha_1)-2\cosh\alpha_1\,n\phi'_n(\alpha_1)+(n-1)\phi'_{n-1}(\alpha_1) = -2b_n \qquad (n \geqq 3)
\end{array}\right\} \quad .\ (36)$$

Writing out equations (35), multiplying by $e^{-n\alpha_1}$ and adding, we have

$$(n+1)\psi'_{n+1}(\alpha_1)e^{-n\alpha_1}-n\psi'_n(\alpha_1)e^{-(n+1)\alpha_1} = 2\sum_{p=0}^{n} a_p e^{-p\alpha_1}$$

or

$$(n+1)\psi'_{n+1}(\alpha_1)-n\psi'_n(\alpha_1)e^{-\alpha_1} = 2e^{n\alpha_1}\sum_{p=0}^{n} a_p e^{-p\alpha_1}. \quad .\ .\ .\ (37)$$

Now, in virtue of (30), we may write the right-hand side of this

$$-2e^{n\alpha_1}\sum_{p=n+1}^{\infty} a_p e^{-p\alpha_1} = -2\sum_{r=1}^{\infty} a_{n+r}e^{-r\alpha_1},$$

and since $\Sigma a_n \cos n\beta$ is supposed convergent this tends to zero as $n$ increases. Hence from (37) we see that the limit of $\psi'_{n+1}(\alpha_1)/\psi'_n(\alpha_1)$ as $n$ increases is $e^{-\alpha_1}$, and hence the functions $\psi'_n(\alpha_1)$ are finite for all values of $n$ and tend to zero as $n$ increases, if the resultant couple acting on $\alpha = \alpha_1$ vanishes.

Multiplying (37) by $e^{n\alpha_1}$ and adding, we have

$$n\psi'_n(\alpha_1)e^{(n-1)\alpha_1} = 2\sum_{q=0}^{n-1}\sum_{p=0}^{q} a_p e^{(2q-p)\alpha_1} = 2\sum_{p=0}^{n-1}\sum_{q=p}^{n-1} a_p e^{(2q-p)\alpha_1},$$

which, on effecting the summation with regard to $q$, leads at once to

$$n\psi'_n(\alpha_1) = 2 \operatorname{cosech} \alpha_1 \sum_{p=0}^{n-1} a_p \sinh(n-p)\alpha_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (38)$$

for $n \geqq 1$.

Treating (34) in a similar way we have

$$n(n+1)(n+2)\psi_{n+1}(\alpha_1)-(n-1)(n)(n+1)\psi_n(\alpha_1)e^{-\alpha_1}$$
$$= 2e^{n\alpha_1}\sum_{p=1}^{n} p\left[d_p+\sinh\alpha_1\psi'_p(\alpha_1)\right]e^{-p\alpha_1}. \quad . \quad . \qquad (39)$$

We can readily show from (35) that

$$\sinh\alpha_1\sum_{p=1}^{n} pe^{-p\alpha_1}\psi'_p(\alpha_1) = \tfrac{1}{2}n(n+1)\left\{\psi'_{n+1}(\alpha_1)-e^{-\alpha}\psi'_n(\alpha_1)\right\}e^{-n\alpha_1}-\sum_{p=1}^{n} pa_pe^{-p\alpha_1},$$

and hence (39) may be written

$$n(n+1)(n+2)\psi_{n+1}(\alpha_1)-(n-1)(n)(n+1)\psi_n(\alpha_1)e^{-\alpha_1} = n(n+1)\left[\psi'_{n+1}(\alpha_1)+e^{-\alpha_1}\psi'_n(\alpha_1)\right]$$
$$+2e^{n\alpha_1}\sum_{p=1}^{n} p(d_p-a_p)e^{-p\alpha_1}.$$

As in the case of $\psi'_n(\alpha_1)$, we can show that the right-hand side tends to zero as $n$ increases if conditions (30) are fulfilled. Hence $\psi_n(\alpha_1)$ is finite for all values of $n$ and tends to zero as $n$ increases, and we have

$$(n-1)(n)(n+1)\psi_n(\alpha_1)e^{(n-1)\alpha_1} = 2\sum_{q=1}^{n-1}\sum_{p=1}^{q} p(d_p-a_p)e^{(2q-p)\alpha_1}$$
$$+\sum_{q=1}^{n-1} q(q+1)e^{q\alpha_1}\left[\psi'_{q+1}(\alpha_1)-e^{-\alpha_1}\psi'_q(\alpha_1)\right],$$

which, on reduction, leads to

$$n(n^2-1)\psi_n(\alpha_1) = 2\operatorname{cosech}\alpha_1\sum_{p=0}^{n-1}\{(n-p)a_p\cosh(n-p)\alpha_1$$
$$+(pd_p-a_p\coth\alpha_1)\sinh(n-p)\alpha_1\} \quad . \quad . \qquad (40)$$

for $n \geqq 2$. Equations (34) do not determine $\psi_1(\alpha_1)$.

From (36) we have

$$2\phi'_2(\alpha_1)e^{-\alpha_1}-\phi'_1(\alpha_1)e^{-2\alpha_1} = \phi'_1(\alpha_1)-2e^{-\alpha_1}(\mathrm{K}\sinh\alpha_1+\mathrm{B}_0\cosh\alpha_1)-2b_1e^{-\alpha_1}, \quad . \qquad (41)$$

and if $n \geqq 2$

$$(n+1)\phi'_{n+1}(\alpha_1)e^{-n\alpha_1}-n\phi'_n(\alpha_1)e^{-(n+1)\alpha_1} = \phi'_1(\alpha_1)-2\mathrm{K}e^{-\alpha_1}\sinh\alpha_1-\mathrm{B}_0$$
$$-2\sum_{p=1}^{n} b_pe^{-p\alpha_1}, \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (42)$$

and hence, if the sequence $\phi_n(\alpha_1)$ is to converge for large values of $n$, we must have

$$\phi'_1(\alpha_1) = \mathrm{B}_0+2\mathrm{K}e^{-\alpha_1}\sinh\alpha_1+2\sum_{p=1}^{\infty} b_pe^{-p\alpha_1}. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (43)$$

From (41) and (42) we have, if $n \geqq 2$,

$$n\phi'_n(\alpha_1)\,e^{(n-1)\alpha_1} = \phi'_1(\alpha_1) - \mathrm{B}_0 + \sum_{q=1}^{n-1}(\phi'_1(\alpha_1) - 2\mathrm{K}e^{-\alpha_1}\sinh\alpha_1 - \mathrm{B}_0)\,e^{2q\alpha_1}$$
$$-2\sum_{q=1}^{n-1}\sum_{p=1}^{q} b_p e^{(2q-p)\alpha_1},$$

from which we obtain

$$n\sinh\alpha_1\,\phi'_n(\alpha_1) = (\phi'_1(\alpha_1) - \mathrm{B}_0)\sinh n\alpha_1 - 2\mathrm{K}\sinh(n-1)\,\alpha_1\sinh\alpha_1$$
$$-2\sum_{p=1}^{n-1} b_p\sinh(n-p)\,\alpha_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (44)$$

for $n \geqq 2$, while $\phi'_1(\alpha_1)$ is given by (43).

Finally we have from (33), omitting the first equation of the series, if $n \geqq 2$,

$$n(n+1)(n+2)\,e^{-n\alpha_1}\phi_{n+1}(\alpha_1) - (n-1)(n)(n+1)\,e^{-(n+1)\alpha_1}\phi_n(\alpha_1)$$
$$= -2(\mathrm{B}_0 + \mathrm{K})\,e^{-\alpha_1}\sinh\alpha_1$$
$$+2\sum_{p=1}^{n} p\,(c_p + \sinh\alpha_1\,\phi'_p(\alpha_1))\,e^{-p\alpha_1}.$$

By the aid of (36) we can reduce the right-hand side to

$$n(n+1)(\phi'_{n+1}(\alpha_1) - e^{-\alpha_1}\phi'_n(\alpha_1))\,e^{-n\alpha_1} + 2\sum_{p=1}^{n} p\,(c_p + b_p)\,e^{-p\alpha_1},$$

from which it appears that $\phi_{n+1}(\alpha_1)$ is finite and tends to zero as $n$ increases, provided that the resultant of the applied forces over $\alpha = \alpha_1$ is zero. We then obtain for $n \geqq 2$,

$$n(n^2-1)\sinh\alpha_1\,\phi_n(\alpha_1) = (\phi'_1(\alpha_1) - \mathrm{B}_0)\,\{n\cosh n\alpha_1 - \coth\alpha_1\sinh n\alpha_1\}$$
$$-\mathrm{K}\,\{(n-1)\sinh n\alpha_1 - (n+1)\sinh(n-2)\,\alpha_1\}$$
$$+2\sum_{p=1}^{n-1}\{(pc_p + b_p\coth\alpha_1)\sinh(n-p)\,\alpha_1$$
$$-(n-p)\,b_p\cosh(n-p)\,\alpha_1\}, \quad . \quad . \quad . \quad . \quad (45)$$

while

$$2\phi_1(\alpha_1) = 2c_0 + \mathrm{B}_0\sinh 2\alpha_1 + \mathrm{K}\,(2\cosh^2\alpha - 1). \quad . \quad . \quad . \quad . \quad . \quad (46)$$

It appears that equations (38), (40), (43), (44), (45) and (46) give the values of $\phi_n(\alpha_1)$, $\psi_n(\alpha_1)$, $\phi'_n$, $(\alpha_1)$, $\psi'_n(\alpha_1)$ for $n \geqq 1$ in terms of $\mathrm{B}_0$, K and the given coefficients $a_n$, &c., with the exception of $\psi_1(\alpha_1)$.

Now we have only assumed $\alpha_1 > 0$ in order to establish the convergence of these functions, and hence the corresponding functions of $\alpha_2$ will be given by the same formulæ with $a'_n$, $b'_n$, $c'_n$, $d'_n$ substituted for $a_n$, $b_n$, $c_n$, $d_n$, provided that the conditions for convergence are satisfied. It may be shown that the new conditions of convergence are identical with (30) and (31), or (30) and (32), according as $\alpha_2 >$ or $< 0$.

The formula for $\phi_1(\alpha_1)$ given in (43), which is itself a condition of convergence, will, however, be replaced by

$$\phi'_1(\alpha_2) = B_0 + 2Ke^{\alpha_2}\sinh\alpha_2 + 2\sum_{n=1}^{\infty} b'_n e^{n\alpha_2} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (47)$$

if $\alpha_2 < 0$.

From (26) we see that the coefficients $A_n$, $B_n$, $C_n$, $D_n$ for $n \geqq 2$ are determined from $\phi_n(\alpha_1)$, $\phi_n(\alpha_2)$, $\phi'_n(\alpha_1)$, $\phi'_n(\alpha_2)$, and similarly $A'_n$, $B'_n$, $C'_n$, $D'_n$ are determined from $\psi_n(\alpha_1)$, $\psi_n(\alpha_2)$, $\psi'_n(\alpha_1)$, $\psi'_n(\alpha_2)$,

The values of $\phi_1(\alpha_1)$, $\phi_1(\alpha_2)$, $\phi'_1(\alpha_1)$, $\phi'_1(\alpha_2)$ will give four equations to determine the three constants $A_1$, $B_1$, $C_1$, and the condition that they shall be consistent gives one relation between $B_0$ and K. The values of $\psi'_1(\alpha_1)$, $\psi'_1(\alpha_2)$ determine the two constants $A'_1$, $C'_1$, and $\psi_1(\alpha_1)$, $\psi_1(\alpha_2)$ are not otherwise determined.

We have thus just sufficient equations to determine the coefficients in (28) with the exception of $B_0$, K, between which we have found one relation. If $\alpha_2 > 0$, so that the region considered lies entirely on one side of the axis $\alpha = 0$, we may take $K = 0$. If on the other hand $\alpha_2 < 0$ the condition that the stress shall vanish at infinity, which is $h\chi \to 0$ when $\alpha$, $\beta \to 0$, gives one more relation between the coefficients, so that in either case $B_0$, K are determined.

We may therefore adopt the following method :—Insert terms of the type (24) or (25) corresponding to the resultant force and couple on each boundary, and calculate the residual stresses over the boundaries. These will now form systems in statical equilibrium over each boundary, and we have shown how to determine an appropriate function of the form (28).

The problem of finding the appropriate stress-function for given tractions over the boundaries might have been approached by investigating the values of $h\chi$ and its normal gradient on the boundaries, on the lines developed by MICHELL.* The direct method which we have adopted is, however, in most cases simpler in our particular co-ordinates.

There is an exception to this rule, namely, when a boundary is free from stress. In this case the boundary conditions assume a very simple form. From (6) we have

$$\frac{\partial}{\partial\alpha}(h\chi) = \text{const} = \rho, \text{ say} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (48)$$

and

$$(\cosh\alpha - \cos\beta)\frac{\partial^2}{\partial\beta^2}(h\chi) - \sin\beta\frac{\partial}{\partial\beta}(h\chi) + \cosh\alpha\,(h\chi) = \rho\sinh\alpha$$

the solution of which is readily found to give

$$h\chi = \rho\tanh\alpha + \sigma(\cosh\alpha\cos\beta - 1) + \tau\sin\beta \quad . \quad . \quad . \quad . \quad . \quad (49)$$

on the boundary considered.

---

* 'Proc. London Mathematical Society,' vol. xxi., 1900, p. 100.

The relations (48) and (49) are the necessary and sufficient conditions that a boundary $\alpha =$ constant should be free from stress. The constants $\rho$, $\sigma$, $\tau$ are MICHELL's three constants of the boundary.

## § 5. A Cylinder or Pipe with Eccentric Bore.

In this section we will consider the problem of a cylinder, whose cross-section is bounded by two non-concentric circles, which is subject to a uniform normal pressure over its internal surface and a different uniform normal pressure over its external surface. By FILON's theorem of generalised plane stress precisely the same analysis will give the average stresses in a plate of the same section under the same applied forces.

Let the boundaries of the cross-section be defined by $\alpha = \alpha_1$ for the internal boundary and $\alpha = \alpha_2$ for the external boundary. Then $\alpha_1$, $\alpha_2$ are positive and $\alpha_1 > \alpha_2$. Let the applied pressures be $P_1$, $P_2$ respectively, so that $\widehat{\alpha\alpha} = -P_1$ on $\alpha = \alpha_1$, $\widehat{\alpha\alpha} = -P_2$ on $\alpha = \alpha_2$ and $\widehat{\alpha\beta} = 0$ on both boundaries.

Let us assume

$$h\chi = B_0\alpha\,(\cosh\alpha - \cos\beta) + (A_1\cosh 2\alpha + B_1 + C_1\sinh 2\alpha)\cos\beta.$$

Calculating $\widehat{\alpha\alpha}$, $\widehat{\beta\beta}$, by means of (6) and applying the boundary conditions, we find the following values for the constants:—

$$B_0 = 2a\,M\,(P_1 - P_2)\cosh(\alpha_1 - \alpha_2)$$
$$A_1 = -aM\,(P_1 - P_2)\sinh(\alpha_1 + \alpha_2)$$
$$C_1 = aM\,(P_1 - P_2)\cosh(\alpha_1 + \alpha_2)$$
$$B_1 = aM\,\{P_1\cosh(\alpha_1 - \alpha_2)\sinh 2\alpha_2 - P_2\cosh(\alpha_1 - \alpha_2)\sinh 2\alpha_1 + (P_1 + P_2)\sinh(\alpha_1 - \alpha_2)\}$$

where, for brevity, we have written

$$M = \tfrac{1}{2}\operatorname{cosech}(\alpha_1 - \alpha_2)\,\{\sinh^2\alpha_1 + \sinh^2\alpha_2\}^{-1}$$

The most important aspect of the problem is the value of the stress $\widehat{\beta\beta}$ in the boundaries, for it is upon this that the strength of the cylinder will depend. This is most readily determined by (7), and we find without difficulty

$$\widehat{\alpha\alpha} - \widehat{\beta\beta} = 4M\,(P_1 - P_2)\,(\cosh\alpha - \cos\beta)\,\{\sinh(\alpha_1 + \alpha_2 - 2\alpha)\cos\beta - \sinh\alpha\cosh(\alpha_1 - \alpha_2)\}$$

so that on $\alpha = \alpha_1$

$$\widehat{\beta\beta} = -P_1 + 4\,(P_1 - P_2)\,M\,(\cosh\alpha_1 - \cos\beta)\{\sinh(\alpha_1 - \alpha_2)\cos\beta + \sinh\alpha_1\cosh(\alpha_1 - \alpha_2)\} \qquad (50)$$

and on $\alpha = \alpha_2$

$$\widehat{\beta\beta} = -P_2 - 4\,(P_1 - P_2)\,M\,(\cosh\alpha_2 - \cos\beta)\{\sinh(\alpha_1 - \alpha_2)\cos\beta - \sinh\alpha_2\cosh(\alpha_1 - \alpha_1)\} \qquad (51)$$

In order to investigate these results further we will consider separately the cases when the cylinder is subject to *either* internal *or* external pressures. There is no greater difficulty in the consideration of the general case, should the necessity arise, except that the formulæ are correspondingly longer.

### *A Cylinder under Internal Pressure.*

If we put $P_2 = 0$, we have on the external surface

$$\widehat{\beta\beta} = -4P_1 M (\cosh \alpha_2 - \cos \beta)\{\sinh (\alpha_1 - \alpha_2) \cos \beta - \sinh \alpha_2 \cosh (\alpha_1 - \alpha_2)\},$$

if $d_1$, $d_2$ denote the distances of the circles $\alpha_1$, $\alpha_2$ from the origin, $r_1$, $r_2$ their radii and $d$ the distance apart of their centres, so that $d = d_2 - d_1$ we may show from (2) that

$$d_1 = a \coth \alpha_1, \qquad d_2 = a \coth \alpha_2$$

$$r_1 = a \operatorname{cosech} \alpha_1, \qquad r_2 = a \operatorname{cosech} \alpha_2$$

and

$$d_1 = (r_2^2 - r_1^2 - d^2)/2d, \qquad d_2 = (r_2^2 - r_1^2 + d^2)/2d$$

$$a^2 = \{r_2^2 - (r_1 + d)^2\}\{r_2^2 - (r_1 - d)^2\}/4d^2.$$

By means of these relations we can reduce the expression for $\widehat{\beta\beta}$ to the form

$$\widehat{\beta\beta} = \frac{2P_1 r_1^2 \{r_2^2 (r_2 - 2d \cos \beta)^2 - (r_1^2 - d^2)^2\}}{(r_1^2 + r_2^2)\{r_2^2 - (r_1 + d)^2\}\{r_2^2 - (r_1 - d)^2\}}.$$

From this and the obvious inequality $d < r_2 - r_1$ we easily see that—

(1) The numerically greatest stress is when $\beta = \pi$, *i.e.*, on the line of centres at the thinnest part of the cylinder. This is always a tension if $P_1$ is positive and is given by

$$\frac{2P_1 r_1^2 (r_2^2 + r_1^2 + 2r_2 d - d^2)}{(r_1^2 + r_2^2)(r_2^2 - r_1^2 - 2r_2 d + d^2)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (52)$$

(2) If the centre distance is greater than half the external radius there is minimum stress at the points corresponding to $\cos \beta = r_2/2d$. This is always negative when $P_1$ is positive and we have maximum compressions equal to

$$\frac{2P_1 r_1^2 (r_1^2 - d^2)^2}{(r_1^2 + r_2^2)\{r_2^2 - (r_1 + d)^2\}\{r_2^2 - (r_1 - d)^2\}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (53)$$

This is always numerically less than the maximum tension. There is a secondary maximum at $\beta = 0$, *i.e.*, on the line of centres at the thickest part of the cylinder, which is equal to

$$\frac{2P_1 (r_2^2 + r_1^2 - 2r_2 d - d^2)}{(r_1^2 + r_2^2)(r_2^2 - r_1^2 + 2r_2 d + d^2)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (54)$$

(3) If the centre distance is less than half the external radius, we have, in addition to the maximum tension (52), a minimum at $\beta = 0$ given by (54). There are no other maxima or minima and the stress decreases steadily from its value at the thinnest part of the cylinder to its value at the thickest part.

On the internal surface we have

$$\widehat{\beta\beta} = -\mathrm{P}_1 + 4\mathrm{P}_1\mathrm{M}\,(\cosh\alpha_1 - \cos\beta)\,\{\sinh(\alpha_1-\alpha_2)\cos\beta + \sinh\alpha_1\cosh(\alpha_1-\alpha_2)\},$$

or, expressed in terms of the radii and centre distance,

$$\widehat{\beta\beta} = -\mathrm{P}_1 + \frac{2\mathrm{P}_1 r_2^2\{(r_2^2-d^2)^2 - r_1^2(r_1+2d\cos\beta)^2\}}{(r_1^2+r_2^2)\{r_2^2-(r_1-d)^2\}\{r_2^2-(r_1+d)^2\}}. \quad . \quad . \quad . \quad . \qquad (55)$$

Hence it may be shown that

(1) If the centre distance is greater than one-half the internal radius the maximum stress in the internal surface occurs at the points corresponding to $\cos\beta = -r_1/2d$ and is

$$-\mathrm{P}_1 + \frac{2\mathrm{P}_1 r_2^2(r_2^2-d^2)^2}{(r_1^2+r_2^2)\{r_2^2-(r_1-d)^2\}\{r_2^2-(r_1+d)^2\}} \quad . \quad . \quad . \quad . \quad . \qquad (56)$$

(2) If the centre distance is less than one-half the internal radius the maximum stress is at $\beta = \pi$, *i.e.*, on the line of centres at the thinnest part of the cylinder. It is

$$-\mathrm{P}_1 + \frac{2\mathrm{P}_1 r_2^2(r_2^2+r_1^2-2r_1d-d^2)}{(r_1^2+r_2^2)(r_2^2-r_1^2-2r_1d-d^2)} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (57)$$

(3) The minimum stress is at $\beta = 0$, the point where the line of centres meets the internal boundary at the thickest part of the cylinder. It is

$$-\mathrm{P}_1 + \frac{2\mathrm{P}_1 r_2^2(r_2^2+r_1^2+2r_1d-d^2)}{(r_1^2+r_2^2)(r_2^2-r_1^2+2r_1d-d^2)}. \quad . \quad . \quad . \quad . \quad . \quad . \qquad (58)$$

This may be shown to be essentially positive if P is positive so that, as would be expected, the internal boundary is everywhere in a state of tension.

### *A Cylinder under External Pressure.*

Putting $\mathrm{P}_1 = 0$ in (50) and (51) we have on the internal surface

$$\begin{aligned}\widehat{\beta\beta} &= -4\mathrm{P}_2\mathrm{M}\,(\cosh\alpha_1-\cos\beta)\,\{\sinh(\alpha_1-\alpha_2)\cos\beta+\sinh\alpha_1\cosh(\alpha_1-\alpha_2)\} \\ &= -\frac{2\mathrm{P}_2 r_2^2\{(r_2^2-d^2)^2-r_1^2(r_1+2d\cos\beta)^2\}}{(r_1^2+r_2^2)\{r_2^2-(r_1-d)^2\}\{r_2^2-(r_1+d)^2\}}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (59)\end{aligned}$$

and on the external surface

$$\begin{aligned}\widehat{\beta\beta} &= -\mathrm{P}_2+4\mathrm{P}_2\mathrm{M}\,(\cosh\alpha_2-\cos\beta)\,\{\sinh(\alpha_1-\alpha_2)\cos\beta-\sinh\alpha_2\cosh(\alpha_1-\alpha_2)\} \\ &= -\mathrm{P}_2 - \frac{2\mathrm{P}_2 r_1^2\{r_2^2(r_2-2d\cos\beta)^2-(r_1^2-d^2)^2\}}{(r_1^2+r_2^2)\{r_2^2-(r_1-d)^2\}\{r^2-(r_1+d)^2\}} \quad . \quad . \quad . \quad . \quad . \quad . \quad . \qquad (60)\end{aligned}$$

Hence if the centre distance is less than half the internal radius the compression in the inner surface decreases steadily from a maximum at the thinnest part of the cylinder to a minimum at the thickest part; otherwise there is a minimum at each of the points and maxima at the points corresponding to cos $\beta = -r_1/2d$. Similarly if the centre distance is less than half the external radius the compression in the outer surface decreases steadily from a maximum at the thinnest part of the cylinder to a minimum at the thickest part; if the centre distance exceeds this value the compression is a maximum at each of these points and minima at the points corresponding to cos $\beta = r_1/2d$.

If in these results we put $d = 0$, we have, for a concentric tube under internal pressure, tensions at the inner and outer surfaces which are respectively

$$\frac{r_2^2+r_1^2}{r_2^2-r_1^2}P_1, \qquad \frac{2r_1^2}{r_2^2-r_1^2}P_1,$$

while for a tube under external pressure the compressions at the inner and outer surfaces are respectively

$$\frac{2r_2^2}{r_2^2-r_1^2}P_2, \qquad \frac{r_2^2+r_1^2}{r_2^2-r_1^2}P_2,$$

These are the well-known formulæ for thick tubes.

## § 6. A SEMI-INFINITE PLATE WITH A CIRCULAR HOLE SUBJECT TO A UNIFORM NORMAL PRESSURE.

If in the results of the last section we put $\alpha_2 = 0$ and $P_2 = 0$, we have the solution for a semi-infinite plate containing a circular hole, which is subject to a uniform normal pressure, and bounded by a straight edge which is free from stress.

We have on the boundary of the hole

$$\widehat{\beta\beta} = -P_1+2P_1 \operatorname{cosech}^2 \alpha_1 (\cosh^2 \alpha_1 - \cos^2 \beta)$$

and on the straight edge

$$\widehat{\beta\beta} = -2P_1 \operatorname{cosech}^2 \alpha_1 (1-\cos \beta) \cos \beta.$$

If $r$ is the radius of the hole, $d$ the perpendicular distance of its centre from the straight edge, and $x$ the distance measured along the straight edge from the foot of the perpendicular,

$$d = a \coth \alpha, \qquad r = a \operatorname{cosech} \alpha_1, \qquad d^2-r^2 = a^2,$$

and

$$x = a \sin \beta/(1-\cos \beta).$$

We have therefore on the straight edge

$$\widehat{\beta\beta} = -4P_1 \frac{r^2(x^2-d^2+r^2)}{(x^2+d^2-r^2)^2}. \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (61)$$

This has a maximum tension at the symmetrical point ($x = 0$) of magnitude

$$4P_1r^2/(d^2-r^2). \qquad (62)$$

At the points $x = \pm\sqrt{(d^2-r^2)}$ it vanishes, and then becomes a compression which reaches a maximum value at points at distances $\pm\sqrt{3}\,(d^2-r^2)$ on either side of the foot of the perpendicular from the centre of the hole, which is numerically equal to one-eighth of maximum tension.

The stress round the circular hole may be represented by a simple geometrical construction. If in fig. 3 the centre of the circular hole is C, Q is any point on the circle, and CA the perpendicular drawn from C to the straight edge, and if Q denote the angle QAC, we easily see that

$$\tan\phi = \sin\beta \operatorname{cosech}\alpha_1,$$

and the stress round the circular hole is

$$\widehat{\beta\beta} = P_1(1+2\tan^2\phi). \qquad (63)$$

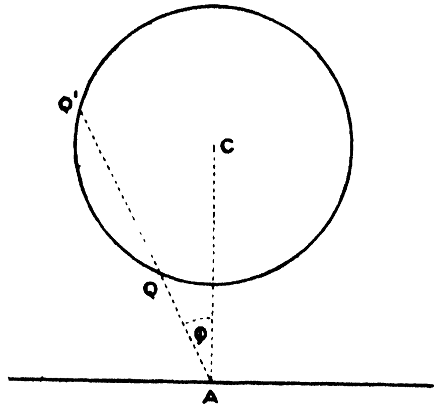


Fig. 3.

Hence the stress is the same at points Q, Q' which lie on the same ray through A. The stress is minimum at the points nearest to and most remote from the straight edge, where it is a tension P numerically equal to the applied pressure. Thus at these points the stress is the same as it would be in the absence of the straight boundary if the plate were infinite. The maximum stresses are at the points of contact of the tangents drawn from A the circular boundary. At these points its value is

$$P_1\frac{d^2+r^2}{d^2-r^2}. \qquad (64)$$

The maximum tension in the circular boundary is equal to the maximum tension in the straight edge if $d = \sqrt{3}r$. In this case each is equal to $2P_1$. If the distance of the hole from the straight edge is greater than this value the maximum tension is at a point on the circular boundary; and if it is less, the maximum stress tension is at the symmetrical point on the straight edge. On the other hand, the point of maximum difference of principal stresses is on the straight edge or the circular boundary, according as $d$ is greater or less than $\sqrt{2}r$. This suggests a simple method of determining whether, for a particular material, rupture occurs at the point of greatest tension or at the point of greatest stress-difference. If a circular hole is bored near the straight edge of a uniform plate, so that the distance of its centre from the edge is greater than $\sqrt{2}$ and less than $\sqrt{3}$ of the radius, and a uniform radial pressure is exerted over the hole in any convenient way and increased until rupture occurs, the crack will begin on the straight edge according to the greatest tension theory, and on the edge of the hole if the greatest stress-difference theory holds.

It will be noted that the stresses produced will become large if the hole is near to the straight edge. The formulæ are so simple that it is hardly worth tabulating their numerical values, but a single example will serve as an illustration. If the shortest distance from the hole to the straight edge is one-tenth of the radius of the hole, the maximum tension in the straight edge is 19·5 times the pressure in the hole.

## §7. A SEMI-INFINITE PLATE CONTAINING AN UNSTRESSED CIRCULAR HOLE AND UNDER A UNIFORM TENSION PARALLEL TO ITS STRAIGHT EDGE.

Let the circular boundary be defined by $\alpha = \alpha_1$, so that if $r$ is its radius and $d$ the distance of its centre from the straight edge,

$$r = a \operatorname{cosech} \alpha_1, \qquad d = a \coth \alpha_1, \qquad d/r = \cosh \alpha_1.$$

At a distance from the hole the stress-function may be taken as $\chi = \frac{1}{2}\mathrm{T}y^2$ where T is the tension, so that, if $\alpha > 0$,

$$\begin{aligned} h\chi_0 &= \tfrac{1}{2}a\mathrm{T} \sinh^2\alpha/(\cosh \alpha - \cos \beta) \\ &= \tfrac{1}{2}a\mathrm{T} \sinh \alpha \left(1 + 2 \sum_{n=1}^{\infty} e^{-n\alpha} \cos n\beta\right) \quad . \quad . \quad . \quad . \quad . \quad . \quad (65) \end{aligned}$$

We have to add to this a stress-function which gives no stress at infinity and no stress over $\alpha = 0$, and is such that the complete stress-function gives no stress over $\alpha = \alpha_1$.

We may omit the term in K in (28), since in this case the region considered lies entirely on one side of $\alpha = 0$, and clearly the required stress-function is even in $\beta$. It may readily be seen that the condition that $\widehat{\alpha\alpha}$ and $\widehat{\alpha\beta}$ shall vanish over $\alpha = 0$ is satisfied by (28) if $\phi_n(0) = 0$ and $\phi'_n(0) = 0$ for $n \geqq 1$, and hence from (26) and (27) $\mathrm{A}_n + \mathrm{B}_n = 0$ and $(n+1)\,\mathrm{C}_n + (n+1)\,\mathrm{D}_n = 0$. We may therefore take for our complete stress-function

$$\begin{aligned} h\chi = a\mathrm{T}\Bigg[\tfrac{1}{2} \sinh \alpha \left\{1 + 2 \sum_{n=1}^{\infty} e^{-n\alpha} \cos n\beta\right\} + \mathrm{B}_0\alpha\,(\cosh \alpha - \cos \beta) + \mathrm{A}_1\,(\cosh 2\alpha - 1) \cos \beta \\ + \sum_{n=2}^{\infty} \left\{ \begin{array}{l} \mathrm{A}_n\,[\cosh(n+1)\,\alpha - \cosh(n-1)\,\alpha] \\ + \mathrm{E}_n\,[(n-1)\sinh(n+1)\,\alpha - (n+1)\sinh(n-1)\,\alpha] \end{array} \right\} \cos n\beta \Bigg] \quad . \quad . \quad (66) \end{aligned}$$

At infinity $\alpha = 0$, $\beta = 0$ the first series diverges, but may of course be replaced by the alternative form in (65). If the second series converges it is clear that at infinity $\chi = \chi_0$.

We must now choose the coefficients in (66) so as to satisfy (48) and (49), and there is no difficulty in finding the following values for the coefficients :—

$$A_1 = \tfrac{1}{2}e^{-2\alpha_1}\operatorname{sech} 2\alpha_1, \qquad B_0 = \operatorname{sech} 2\alpha_1 \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad . \quad (67)$$

$$A_n = -\frac{n^2 \sinh^2\alpha_1 - n\sinh\alpha_1\cosh\alpha_1 + e^{-n\alpha_1}\sinh n\alpha_1}{2\{\sinh^2 n\alpha_1 - n^2\sinh^2\alpha_1\}} \quad . \quad . \quad . \quad . \quad (68)$$

$$E_n = \frac{n\sinh^2\alpha_1}{2\{\sinh^2 n\alpha_1 - n^2\sinh^2\alpha_1\}}.$$

Substituting in (66) we have for the complete stress-function,

$$h\chi = \alpha\mathrm{T}\left[\begin{array}{l}\alpha\operatorname{sech} 2\alpha_1(\cosh\alpha-\cos\beta)+\tfrac{1}{2}\sinh\alpha+\operatorname{sech} 2\alpha_1\cosh(2\alpha_1-\alpha)\sinh\alpha\cos\beta \\ +\displaystyle\sum_{n=2}^{\infty}\frac{\left\{\begin{array}{l} n\sinh\alpha_1\sinh(\alpha-\alpha_1)\sinh n\alpha \\ \qquad -\sinh\alpha\sinh n(\alpha+\alpha_1)\sinh n\alpha_1\end{array}\right\}\cos n\beta}{\sinh^2 n\alpha_1-n^2\sinh^2\alpha_1}\end{array}\right]. \quad (69)$$

We may now calculate the stress $\widehat{\beta\beta}$ in the boundaries by means of (6). We find, on the circular boundary $\alpha = \alpha_1$,

$$\widehat{\beta\beta}_1 = 2\mathrm{T}(\cosh\alpha_1-\cos\beta)\left\{\sinh\alpha_1\operatorname{sech} 2\alpha_1 + \sum_{n=2}^{\infty}\mathrm{M}_n\cos n\beta\right\} \quad . \quad . \quad . \quad (70)$$

where

$$\mathrm{M}_n = \frac{n(n-1)\sinh(n+1)\alpha_1 - n(n+1)\sinh(n-1)\alpha_1}{2\{\sinh^2 n\alpha_1 - n^2\sinh^2\alpha_1\}} \quad . \quad . \quad . \quad . \quad (71)$$

The stress in the straight boundary cannot be directly determined from (69), for it is found that the resulting series diverges for $\alpha = 0$. We can, however, find without difficulty from (66) that when $\alpha = 0$,

$$\widehat{\beta\beta}_0 = \mathrm{T}\left\{1+(1-\cos\beta)\sum_{1}^{\infty}\mathrm{P}_n\cos n\beta\right\} \quad . \quad . \quad . \quad . \quad . \quad . \quad (72)$$

where $\mathrm{P}_n = 4n\mathrm{A}_n$.

The series in (70) converges only slowly, unless $\alpha_1$ is large, and for convenience in computation we may transform it by separating the more slowly converging part.

Let

$$\mathrm{M}_n = 2n(n\sinh\alpha_1 - \cosh\alpha_1)e^{-n\alpha_1} + \mathrm{N}_n, \quad . \quad . \quad . \quad . \quad . \quad . \quad (73)$$

and we readily obtain

$$2(\cosh\alpha_1-\cos\beta)\sum_{n=1}^{\infty}n(n\sinh\alpha_1-\cosh\alpha_1)e^{-n\alpha_1}\cos n\beta = 1-\frac{2\sinh^2\alpha_1\sin^2\beta}{(\cosh\alpha_1-\cos\beta)^2},$$

Substituting in (70) we have

$$\widehat{\beta\beta}_1 = 2\mathrm{T}\left\{1-\frac{2\sinh^2\alpha_1\sin^2\beta}{(\cosh\alpha_1-\cos\beta)^2}\right\}$$

$$+2\mathrm{T}(\cosh\alpha_1-\cos\beta)\left\{\sinh\alpha_1\operatorname{sech} 2\alpha_1+2e^{-2\alpha_1}\cos\beta+\sum_{n=2}^{\infty}\mathrm{N}_n\cos n\beta\right\}. \quad (74)$$

If $\theta$ is the angle between the radius to the point $\alpha_1$, $\beta$ and the perpendicular to the straight edge, then

$$\sin\theta = \frac{\sinh\alpha_1 \sin\beta}{\cosh\alpha_1 - \cos\beta},$$

and if $\alpha_1$ is large (74) reduces to $\widehat{\beta\beta_1} = \mathrm{T}(1+2\cos 2\theta)$, which agrees with the known result for a hole in an infinite plate and gives compression numerically equal to T at the extremities of the diameter parallel to the tension, and tensions equal to 3T at the extremities of the perpendicular diameter.

The numerical values of the coefficients $\mathrm{P}_n$, $\mathrm{N}_n$ are given in Tables I. and II. respectively. It will be noted from Table I. (that, as $\alpha_1$ increases, $\mathrm{P}_2$ tends to become

TABLE I.

| $\alpha_1$. | 0·6. | 0·8. | 1·0. | 1·2. | 1·4. | 1·6. | 1·8. | 2·0. | 2·2. | 2·4. |
|---|---|---|---|---|---|---|---|---|---|---|
| $\mathrm{P}_1$ | 0·3327 | 0·1567 | 0·0719 | 0·0327 | 0·0147 | 0·0066 | 0·0030 | 0·0013 | 0·0006 | 0·0003 |
| $-\mathrm{P}_2$ | 3·5861 | 2·0401 | 1·2545 | 0·7987 | 0·5180 | 0·3400 | 0·2247 | 0·1493 | 0·0994 | 0·0664 |
| $-\mathrm{P}_3$ | 2·2393 | 1·0622 | 0·5110 | 0·2448 | 0·1160 | 0·0543 | 0·0251 | 0·0115 | 0·0053 | 0·0024 |
| $-\mathrm{P}_4$ | 1·3557 | 0·4874 | 0·1699 | 0·0570 | 0·0185 | 0·0059 | 0·0018 | 0·0006 | 0·0002 | 0·0001 |
| $-\mathrm{P}_5$ | 0·7602 | 0·1970 | 0·0474 | 0·0108 | 0·0024 | 0·0005 | 0·0001 | | | |
| $-\mathrm{P}_6$ | 0·3964 | 0·0713 | 0·0116 | 0·0018 | 0·0003 | | | | | |
| $-\mathrm{P}_7$ | 0·1934 | 0·0237 | 0·0026 | 0·0003 | | | | | | |
| $-\mathrm{P}_8$ | 0·0891 | 0·0073 | 3·0005 | | | | | | | |
| $-\mathrm{P}_9$ | 0·0391 | 0·0022 | 0·0001 | | | | | | | |
| $-\mathrm{P}_{10}$ | 0·0165 | 0·0005 | | | | | | | | |
| $-\mathrm{P}_{11}$ | 0·0067 | 0·0002 | | | | | | | | |
| $-\mathrm{P}_{12}$ | 0·0027 | | | | | | | | | |
| $-\mathrm{P}_{13}$ | 0·0010 | | | | | | | | | |
| $-\mathrm{P}_{14}$ | 0·0004 | | | | | | | | | |
| $-\mathrm{P}_{15}$ | 0·0001 | | | | | | | | | |
| $-\mathrm{P}_{16}$ | 0·0001 | | | | | | | | | |

TABLE II.

| $\alpha_1$. | 0·6. | 0·8. | 1·0. | 1·2. | 1·4. | 1·6. | 1·8. | 2·0. | 2·2. | 2·4. |
|---|---|---|---|---|---|---|---|---|---|---|
| $\mathrm{N}_2$ | 1·4649 | 0·7716 | 0·4139 | 0·2240 | 0·1219 | 0·0665 | 0·0364 | 0·0199 | 0·0109 | 0·0060 |
| $\mathrm{N}_3$ | 0·7457 | 0·2647 | 0·0914 | 0·0306 | 0·0100 | 0·0032 | 0·0010 | 0·0003 | 0·0001 | |
| $\mathrm{N}_4$ | 0·3238 | 0·0719 | 0·0148 | 0·0029 | 0·0005 | 0·0001 | | | | |
| $\mathrm{N}_5$ | 0·1232 | 0·0162 | 0·0019 | 0·0002 | | | | | | |
| $\mathrm{N}_6$ | 0·0421 | 0·0032 | 0·0002 | | | | | | | |
| $\mathrm{N}_7$ | 0·0131 | 0·0005 | | | | | | | | |
| $\mathrm{N}_8$ | 0·0038 | 0·0001 | | | | | | | | |
| $\mathrm{N}_9$ | 0·0010 | | | | | | | | | |
| $\mathrm{N}_{10}$ | 0·0003 | | | | | | | | | |
| $\mathrm{N}_{11}$ | 0·0001 | | | | | | | | | |

large compared to the other coefficients. Hence, when the hole is at a considerable distance from the straight edge, the stress in the straight edge approximates to

$$T\{1-C\cos 2\beta(1-\cos\beta)\}$$

where C is a small positive constant.

This shows that the stress in the straight edge is a minimum at the mid-point, increases to a maximum as we move outwards, then diminishes to a second minimum, and finally increases steadily to the value T at infinity, where $\beta = 0$.

In fig. 4 we have plotted the graphs of the stresses in the boundaries for a case in which the hole is fairly near to the straight edge, $\alpha_1 = 0\cdot8$, for which the shortest distance between the two boundaries is approximately one-third of the radius of the circle. It will be noted that the general character of the stresses is not affected by the proximity of the straight edge. It will be remembered that when the hole is at a great distance from the straight edge there are maximum stresses of 3T at the extremities of the diameter perpendicular to the straight edge, with points of maximum compression numerically equal to T lying between. For $\alpha_1 = 0\cdot8$ we find that the maxima occur at the same places but are increased, the increase being more marked at the point nearest to the straight edge, where the tension is 4·366T, while its value at the point most remote from the straight edge is 3·266T. The stress in the straight edge also maintains the same general character as it exhibits when the hole is at a great distance

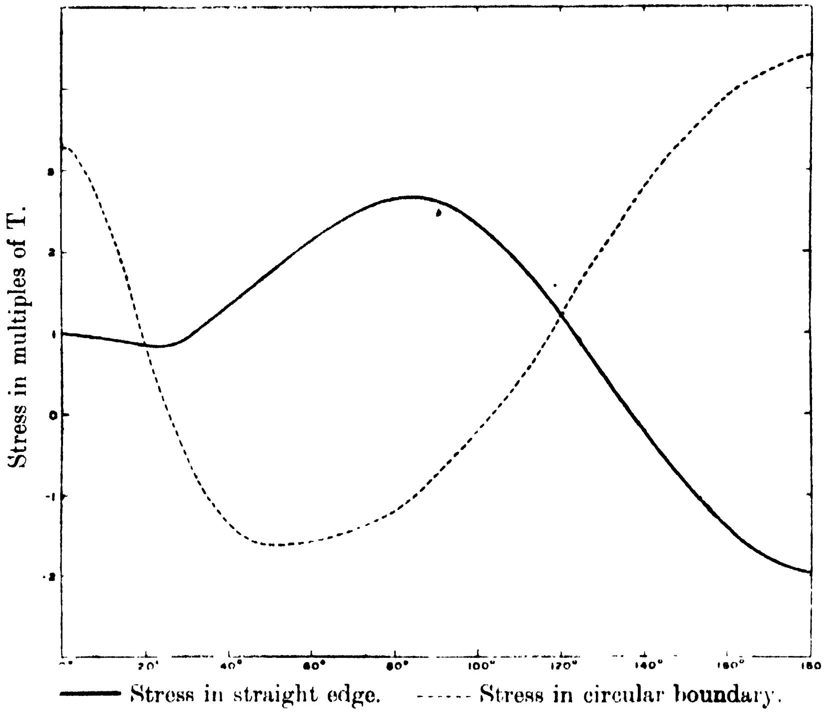


Fig. 4.

from the straight edge. Here again, however, the maxima and minima are accentuated. The minimum at the central point has decreased and has become a compression numerically equal to 1·956T.

It appears that for the range of values which we have investigated the maximum stress is on the circular boundary at the point of nearest approach to the straight edge. Its value for different values of the ratio of the distance of the centre of the hole from the straight edge to the radius of the hole, together with the stresses at the centre of the straight edge and at the most remote point of the circular boundary, is shown in Table III.

TABLE III.

| $\alpha_1$. | Ratio of distance of centre from edge to radius of hole. | Stress at mid-point of straight edge | Stress at nearest point of circular boundary. | Stress at most remote point of circular boundary. |
|---|---|---|---|---|
| 0·6 | 1·185 | −4·080T | 5·064T | 3·362T |
| 0·8 | 1·337 | −1·956 | 4·366 | 3·266 |
| 1·0 | 1·543 | −0·895 | 3·919 | 3·201 |
| 1·2 | 1·811 | 0·269 | 3·609 | 3·152 |
| 1·4 | 2·151 | +0·134 | 3·396 | 3·115 |
| 1·6 | 2·577 | 0·405 | 3·254 | 3·087 |
| 1·8 | 3·107 | 0·591 | 3·162 | 3·065 |
| 2·0 | 3·762 | 0·721 | 3·103 | 3·048 |
| 2·2 | 4·568 | 0·810 | 3·065 | 3·035 |
| 2·4 | 5·557 | 0·871 | 3·043 | 3·025 |
| ∞ | ∞ | 1·000 | 3·000 | 3·000 |

It will be noted that when the hole is very near to the straight edge, so that the two boundaries are separated only by a narrow connecting piece, the stress in this piece consists of a very large tension on the inside and a numerically slightly less compression on the outside. Hence, as might be expected from general considerations, the stress in this narrow connecting piece is a bending moment accompanied by a certain amount of tension.

These results may be compared with some experimental results recently obtained by Prof. COKER and Messrs. K. C. CHAKKO and Y. SATAKE* by optical means. These deal with the stresses in a strip of *finite* width under tension with a circular hole centrally placed, whereas we have considered the case of a semi-infinite plate with a circular hole near its straight edge. The problems are therefore not quite comparable; but as in each case the critical region will clearly be near the minimum section between the hole and a straight boundary, the two problems may be expected to exhibit the same general characteristics. For the strip of finite width it is found that there is maximum stress in the circular boundary at the points of nearest approach to the straight edges and minimum stress at the points of the straight edge immediately opposite the centre of the hole. Moreover as the radius of the hole is increased in proportion to the distance of its centre from the edges of the strip these maxima and minima become more pronounced. In all the cases examined experimentally the minimum stress in the straight edge remains a tension, but Prof. COKER surmises that if the radius of the hole were still further increased in proportion to the width of the strip this minimum stress would become a compression. All these results agree qualitatively with the theoretical results established in this paper for the semi-infinite plate, and allowing for the difference in the two problems they may be taken as a substantial experimental verification.

---

* 'Transactions of the Institution of Engineers and Shipbuilders in Scotland,' vol. lxiii., Part I., p. 33, 1919.

# X. *The Aerodynamics of a Spinning Shell.*

*By* R. H. FOWLER, E. G. GALLOP, C. N. H. LOCK *and* H. W. RICHMOND, *F.R.S.*

Received January 14,—Read May 6, 1920.

## CONTENTS.

Page
Introduction . . . . . . . . . . 296

PART I.—A GENERAL DESCRIPTION OF THE MOTION OF A SPINNING SHELL AND THE PRINCIPAL EXPERIMENTAL RESULTS.

§ 1.0. The classical theory of the plane trajectory . . . . . . . . 299
§ 1.1. The detailed specification of the complete force system . . . . . . . . 301
§ 1.2. The numerical results of the experiments . . . . . . . . 307
§ 1.3. A description in general terms of the angular motion of the axis of a shell . . . . . . 311

PART II.—DETAILS OF THE EXPERIMENTAL ARRANGEMENTS AND MATERIAL.

§ 2.0. General arrangements . . . . . . . . 315
§ 2.1. Measurement of the holes in the cards . . . . . . . . 319
§ 2.2. Determination of the dynamical constants . . . . . . . . 323
§ 2.3. Possible disturbing factors . . . . . . . . 324

PART III.—METHODS OF OBTAINING AND SOLVING THE EQUATIONS OF MOTION OF A SPINNING SHELL.

§ 3.0. Introductory . . . . . . . . 325
§ 3.1. The general vector equations of motion . . . . . . . . 326
§ 3.2. Equations of motion of type $\alpha$ . . . . . . . . 329
§ 3.3. Equations of motion of type $\beta$ . . . . . . . . 331
§ 3.4. Equations of motion of type $\gamma$ . . . . . . . . 333
§ 3.5. The additional force components H, J and K . . . . . . . . 334
§ 3.6. The approximate solution of equations of type $\alpha$ . . . . . . . . 336
§ 3.7. The solution of equations of type $\gamma$ . . . . . . . . 343

PART IV.—ANALYSIS OF THE EXPERIMENTAL RESULTS.

§ 4.0. Equations of motion in polar co-ordinates . . . . . . . . 344
§ 4.1. Analysis of the experimental results . . . . . . . . 347
§ 4.2. Determination of the motion of the shell in space . . . . . . . . 357
§ 4.3. Estimate of the errors in the various solutions . . . . . . . . 362

PART V.—SUMMARY AND CONCLUSION.

§ 5.0. Summary of preceding results . . . . . . . . 367
§ 5.1. Problems for further discussion . . . . . . . . 368
§ 5.2. Effect of size and shape of shell . . . . . . . . 369

[Published December 3, 1920.

## INTRODUCTION.

THIS paper contains the results, theoretical and experimental, of work undertaken, at the request of the Ordnance Committee, by the authors as Technical Officers of the Munitions Inventions Department. Permission to publish such parts as appear to be of general scientific interest has now been granted by the Ordnance Committee and the Director of Artillery. The publication of this paper has received their sanction.

The experiments in question were carried out at the firing ground of H.M.S. "Excellent," Portsmouth; the Experimental Department, H.M.S. "Excellent," also provided the 3-inch guns used and the material for the construction of the range. The authors' best thanks are due to the officers of this department, especially Lieut.-Commander R. F. P. MATON, O.B.E., R.N., without whose cordial co-operation these experiments could never have been carried out; also to the other officers of the Munitions Inventions Department who assisted in the heavy work of making and analysing the observations. The aeronautical measurements at low velocities, required for comparison, were made in the wind channels of the National Physical Laboratory, by arrangement with the Director and the Superintendent of the Aeronautical Department, to whom also we wish to express our thanks.

The subject of this paper is the motion of a spinning shell through air at velocities both greater and less than the velocity of sound. We first attempt to describe the motion of the spinning shell, considered as a rigid body, under the effects of gravity and the reaction of the air; this latter is supposed to be known in terms of the position and velocity co-ordinates of the shell, and the state of the air through which it moves. We are thus concerned throughout with the "aerodynamical" problem of the motion of the shell alone, and not with the general "hydrodynamical" problem of the motion of the complete system formed by the shell and air together. The motion of the shell thus described is then compared with the results of experiments, and the more important components of the force system imposed by the air are determined numerically as functions of certain variables such as the velocity of the centre of gravity of the shell. The actual experiments consist of observations of the initial motion of the shell (more particularly the angular motion of its axis of symmetry), over a limited range near the muzzle of the gun. The velocities experimented with range from 40 f.s.* to 2300 f.s., that is from about 0·04 to 2·1 times the velocity of sound. Using the values of the components so determined, the actual motion of the shell can be calculated with equal certainty in the more general cases which are inaccessible to direct and detailed observation.

As stated above, we make no attempt to attack the hydrodynamical problem. Such an attack is probably not yet feasible. By obtaining, however, an accurate descriptive knowledge of the force system imposed by the air, and the allied system of pressure distribution over the surface of the shell, material is provided on which a successful attack on the hydrodynamical problem may some day be based. A first contribution to a knowledge of the force system is made by the present paper. It is hoped to make a similar contribution to a knowledge of the pressure distribution in another place.*

---

* This velocity was obtained in a wind channel, using a current of air and stationary shell. The lowest velocity used in actual firing experiments was 880 f.s.

The problem proposed for discussion is of course by no means novel.† In the earlier work which is summarised by CRANZ (*cf.* (4)) the treatment of the equations of motion is often open to criticism, in view of the lack of sufficient justification for the necessary simplifying approximations. The classical theoretical results, such as MAYEVSKI's equation for the *drift*‡ (see § 4.2, equation (4.204)) have therefore hitherto justly commanded little confidence. The discussion, moreover, is of necessity based on *a priori* assumptions as to the nature of the complete force system. Unless the results of these assumptions are brought to the test of detailed experiment, the assumptions themselves must remain unjustified and unjustifiable. It should be stated here that the theory and experiments described in this paper confirm the classical theoretical results. CRANZ's own experiments (*cf.* (5)) were expressly designed to explode the fallacy that the axis of the shell, in steady motion, *precesses right round* the direction of motion of its centre of gravity. In this they are successful, but they were only carried out at low velocities, and give little in the way of quantitative results. The only real comparison of theory with experiment, which has hitherto been made, is the comparison of the observed and calculated§ values of the drift. But the observed drift is the integrated result of the disturbing forces over a considerable arc of the trajectory, and moreover, can only be disentangled with difficulty from the effect of any cross wind that may be blowing. The observed drift does not therefore serve to determine the force system with any success, though it may be used to check the values of the components otherwise determined (§ 4.21).

It may, therefore, be stated in general terms that, up to the present, there is no knowledge of the force system acting on any shell at high velocities, except when the shell is moving "nose on," *i.e.*, when its axis of symmetry and the direction of motion of its centre of gravity coincide.*

---

* "The Pressure Distribution on the Head of a Shell Moving at High Velocities," 'Roy. Soc. Proc.,' A, Vol. XCVII., p. 202.

† See for example :—

P. CHARBONNIER. (1) 'Traité de Balistique Extérieure,' ed. 2, Bk. V., Ch. IV. (2) 'Balistique Extérieure Rationnelle,' vol. II., Ch. IX.

C. CRANZ. (3) 'Lehrbuch der Ballistik; Aeussere Ballistik,' 1917, Ch. X. (4) 'Encyklopädie der Mathematischen Wissenschaften,' vol. IV., Part II., p. 185, Art. 18, "Ballistik." (5) 'Zeitschrift für Mathematik und Physik,' vol. XLIII., pp. 133, 169.

J. PRESCOTT. (6) 'Phil. Mag.,' Ser. 6, vol. XXXIV., p. 332.

Further references to previous authors will be found in (4), and the best account of CRANZ's own work in (5).

‡ The lateral departure of the projectile from the vertical plane containing the initial tangent to the path of the centre of gravity of the shell.

§ Actually, also, the important term in the calculated drift depends only on the ratio of two components of the force system, and not on their absolute values.

CRANZ (*cf.* (5)) and CHARBONNIER (*cf.* (2)) make little progress in the treatment of the general equations of motion. PRESCOTT (*cf.* (6)) makes an appreciable advance in the reductions of the general equations of motion to a tractable form, which is not too restricted in application, and gives an exact solution of his reduced equations in the simple case in which all the components of the impressed force system vary as the square of the velocity of the shell. We understand, also, that the problem of the initial motion of the shell has been recently treated by M. ESCLANGON and M. GARNIER, of the French Artillerie de Marine, with results that are closely analogous to ours, but we have not seen their work.

We therefore propose, in this paper, to give in Part III. a detailed account of the complete equations of motion of a spinning shell, moving through air, and to justify as far as possible the reduction of these equations to various useful approximate forms, some of which are classical. To do this, it is of course necessary to start from certain *a priori* assumptions as to the nature of the complete force system. These assumptions, which are far less restrictive than any that have hitherto been used, are carefully analysed when they are introduced. We then, in Part IV., submit the theoretical results so obtained to the test of the experiment described in Part II.; we are thus able to justify to some extent our *a priori* assumptions, and to obtain numerical results of some precision as to the more important components of the force system acting on the shell, in the general case. These numerical results, with a general description of the actual motion of a shell, will be found in Part I.

We have seen that the information to be obtained by comparison of the observed and calculated values of the drift is of very limited value. Two alternative methods are available, both of which are employed in this paper:—

(1) The complete force system on a model shell at rest in a uniform current of air may be determined by observations in a wind channel.†

(2) Certain components of the force system on a shell moving at high velocity may be deduced from the measurements of its oscillations just after leaving the muzzle.

The highest velocity obtainable at present by the first method is 80 f.s., but by means of the "square law" (see § 1.01) the results may be extended to velocities as great as 700 f.s. For higher velocities it is necessary to fall back on the second method which is the principal subject of this paper.

---

* In this case the force system has only one component of practical importance, namely, the resistance of the air, acting in the opposite direction to the relative motion of air and shell. This force component is here called *the drag*, in conformity with aerodynamical usage. The numerical values of the drag are known with fair accuracy for certain external shapes of shell and ordinary atmospheric conditions.

† For a full description of the construction of the wind channels at the National Physical Laboratory, and their use in measuring forces on model aircraft, see COWLEY and LEVY, "Aeronautics in Theory and Experiment."

For this purpose the shell is fired horizontally through a series of cards such as are used for measuring the jump* of the gun on firing. From the shape of the holes in the cards the actual motion of the axis of the shell can be reconstructed. Initial disturbances at the muzzle give rise to angular oscillations of the shell of sufficient amplitude for accurate measurement. These oscillations are very similar to those of the axis of a spinning top under gravity. If, as a first approximation, we regard the centre of gravity of the shell as constrained to move uniformly in a straight line over the range containing the cards, and ignore frictional damping forces in both cases, then the angular motion of the axis of the top and the axis of the shell are identical, provided that (1) the top and shell have the same axial spin and axial moment of inertia; (2) the transverse moment of inertia of the top about its point of support is equal to the transverse moment of inertia of the shell about its centre of gravity; and (3) the moment of gravity about the point of the top is equal to the moment of the force system on the shell about its centre of gravity.

In this approximate case the formal solution of the two problems is identical. As is explained in § 1.3, from the periods of the oscillations of the axis of the top or shell, we can deduce the moment of the disturbing couple and *vice versa*. In the same way the nature of the decay of the oscillations can be used to determine the damping forces.

In conclusion, we feel that a word of apology may be needed for the length of the introductory part of this paper. We do not here emphasise the applications to practical gunnery of the results obtained, but these are of some importance. We have, therefore, thought it desirable that the results should be presented in such a form as to be available to those who are concerned with the practical results, but who are not prepared to follow in detail the arguments of Parts III. and IV. At the same time it has been necessary to avoid statements which, without explanations, might convey little meaning to those who have not been technically concerned with ballistics and aerodynamics. It does not appear possible to achieve these objects except at the expense of a somewhat lengthy Introduction and Part I.

## PART I.—A GENERAL DESCRIPTION OF THE MOTION OF A SPINNING SHELL AND THE PRINCIPAL EXPERIMENTAL RESULTS.

### § 1.0. *The Classical Theory of the Plane Trajectory.*

According to the classical theory, a shell is supposed to move in a resisting medium like a particle on which the only forces acting are gravity, and a resistance tangential to its path, depending only on the velocity of the particle and the state of the undisturbed medium. In such circumstances the path of the particle lies in a vertical plane and is called *the plane trajectory.** This theory would be exact for a shell if the axis of the shell always pointed along the tangent to the path of its centre of gravity. The total reaction between the air and the shell would then, as required, take the form of a single force, called the drag, acting by symmetry tangentially to the path of the centre of gravity, and depending only on the velocity and shape of the shell and the state of the medium. The equations of motion resulting in this simple case are insoluble in finite terms for the actual law of resistance of the air; in practice they are capable of rapid numerical solution to any desired degree of accuracy, by a variety of methods of step-by-step integration, when the drag has been specified with corresponding accuracy.

---

* The angle between the axis of the bore before firing and the initial tangent to the path of the centre of gravity of the shell.

In order to specify the drag completely it is necessary to consider with some care what are the variables on which the drag for a given shell can depend to an appreciable extent. This question is, as yet, by no means settled, and a few of the more important considerations are summarised in § 1.01. This fact does not concern us here to a very serious extent; an incomplete specification of the variables on which the drag (or, in the general case, the complete force system) depends will only invalidate the results of observation when an attempt is made to apply them to widely different conditions of the state of the resisting medium, or of the motion of the shell. The validity is unaffected when the experimental conditions are approximately repeated. It may be assumed that, in this case of symmetry, a fairly adequate expression for the drag is given by the equation

$$(1.001) \qquad \mathrm{R} = \rho v^2 r^2 f_{\mathrm{R}}(v/a),$$

where R is the total drag, $\rho$ the density of the air (or other medium), $r$ the radius of the shell, $v$ the velocity of the shell, and $a$ the velocity of sound in the undisturbed medium; all these quantities, of course, are to be measured in a consistent set of units. In the numerical work in this paper the foot, pound, second system will be used.

Since $\rho v^2 r^2$ has the dimensions of a force, the function $f_{\mathrm{R}}$ is a numerical coefficient, independent of the system of units chosen, called the *drag coefficient.* Existing determinations of this coefficient as a function of $v/a$ are very inadequate from a scientific point of view; satisfactory ones could now be made. We shall not be concerned here with the determination of this coefficient, whose value we shall only require roughly in the analysis of our experiments. We may therefore regard $f_{\mathrm{R}}$ as known for all values of the argument from 0 to 3, for shells of the particular external shapes which we use, moving through dry (or not too nearly saturated) air, whose temperature is not too widely different from 0° C.

---

* From the point of view of this paper, we regard the whole theory of the plane trajectory as "classical," though its adequate treatment was only evolved during the last years of the war.

1.01. *The Functional Form of the Drag Coefficient.*—A careful consideration of the possible forms of the function $f_R$, from the points of view of the kinetic theory of gases and the theory of dimensions, suggests that $\gamma$, $l/r$, and $\sigma/r$ should be possible arguments of $f_R$, besides $v/a$. Here $\gamma$ is the ratio of the specific heats of the gas, $l$ is the mean free path, and $\sigma$ the effective diameter of its molecules. We may, if desired, replace $l/r$ by the more usual viscosity argument $vr/\nu$, where $\nu$ is the kinematical coefficient of viscosity. Wind channel work on aerofoil and airscrew models shows that the argument $vr/\nu$ is of great importance at low velocities. Its effects, however, in the case of shell models seem almost to have disappeared by the time a velocity of 40 f.s. (or at any rate 75 f.s.) is reached. RAYLEIGH* obtains formulæ for the pressure on a piston moving in a pipe, which show the kind of way in which $\gamma$, as well as $v/a$, might enter into the expression for $f_R$. Variations of $\gamma$ are, however, very small in practice. There is experimental evidence that some argument, other than $v/a$ or $\gamma$, has an appreciable effect in practice, and that this argument is probably not the viscosity term in the ordinary sense. It is not possible to pursue the question further here, or to assemble in detail the evidence, which is to be found in various minutes of the Ordnance Committee.

So long as the stream lines of the flow remain unaltered by a change of velocity, the motion remains dynamically similar, the drag varies as $v^2$, and the coefficient $f_R$ must be a constant. The drag is then said to obey *the square law*. Experiments with air screws, of high peripheral speed, appear to show that, up to values of $v/a$ as great as 0·7, there is no serious departure from the square law once a certain minimum velocity is exceeded, above which the ordinary viscosity effects become unimportant; this appears, from all the evidence, to be the case also for shells, the minimum velocity being of the order of 50 f.s. As velocities of less than 100 f.s. may be ignored in ballistics, it is therefore customary to assume that the drag obeys the square law exactly for all velocities less than about 0·7$a$. For all such velocities the stream lines of the flow will remain nearly unaltered and the motion will be dynamically similar.

Above this velocity (0·7$a$.) the effects of the compressibility of the air become rapidly of great importance, and the whole nature of the air-flow changes as $a$, the velocity of sound, is reached and exceeded. These effects are represented by the variation of $f_R$ as a function of $v/a$. A good typical curve showing this variation is given by CRANZ.† Another example will be found in fig. 4.

We have so far ignored the fact that the shell is actually spinning about its axis of symmetry. There is no evidence to show that the drag, in the case of symmetry, is appreciably affected by the spin, and its neglect is probably justified.

A more important question is the legitimacy of assuming, as we have tacitly done in (1.001), that the drag does not depend appreciably on the acceleration of the shell. With regard to the acceleration at low velocities, it is known that the effect of the air is to increase the virtual mass of any body by an amount of the order of the mass of air displaced. This is an increase of the order of 1 part in 2000, and is entirely negligible. At higher velocities, and on the general question, direct experimental evidence is unfortunately lacking. It is, however, difficult to see, by theoretical reasoning, how the past history of the shell can have any *large* effect, and there is sufficient general experimental evidence that (1.001) is, on the average, an adequate representation of the drag in the case of symmetry to be certain that the past history is of little importance, except conceivably for a very limited range of velocities, for example, in the neighbourhood of $a$, the velocity of sound.

## § 1.1. *The Detailed Specification of the Complete Force System.*

The theory discussed in this paper treats the shell as a rigid body which is a solid of revolution, so that its axis of symmetry coincides with a principal axis of inertia.

---

[*] "Aerial Plane Waves of Finite Amplitude," 'Scientific Papers,' vol. V., or 'Roy. Soc. Proc.,' A, vol. LXXXIV. See in particular the last section of the paper.

[†] 'Encyklop. der Math. Wiss.,' vol. IV., Part II., p. 197.

It aims at determining the exact angular motion, as well as the motion of the centre of gravity. It confirms the classical theory of the plane trajectory (in accordance with the results of experiment), by showing that the divergences of the axis of the shell from the tangent to its path are generally small, but it aims, further, at determining the magnitude and effect of these divergences.

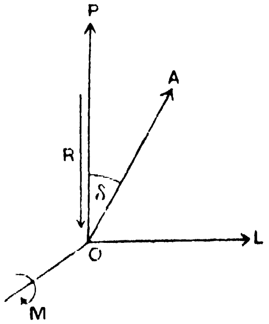


Fig. 1.

In this general case the force system to be specified is more elaborate than in the case of the classical theory. In accordance with aerodynamical usage, we call the angle between the axis of symmetry of the shell and the direction of motion of its centre of gravity *the yaw*, and denote it by $\delta$. When the shell, regarded for the moment as without axial spin, has a yaw $\delta$, and the axis of the shell OA and the direction of motion OP remain in the same relative positions, the force system can by symmetry be represented, as shown in fig. 1, by the following components, specified according to aerodynamical usage.

(1) *The drag*, R, acting through the centre of gravity O, in the direction of motion OP reversed.

(2) A component L, at right angles to R, called *the cross wind force*, which acts through O in the plane of yaw POA, and is positive when it tends to move O in the direction from P to A.

(3) *A moment* M about O, which acts in the plane of yaw, and is positive when it tends to increase the yaw.

By analogy with § 1.0, we assume the following forms for R, L, and M :—

$$\text{(1.101)} \qquad R = \rho v^2 r^2 f_R(v/a, \delta),$$

$$\text{(1.102)} \qquad L = \rho v^2 r^2 \sin\delta f_L(v/a, \delta),$$

$$\text{(1.103)} \qquad M = \rho v^2 r^3 \sin\delta f_M(v/a, \delta).$$

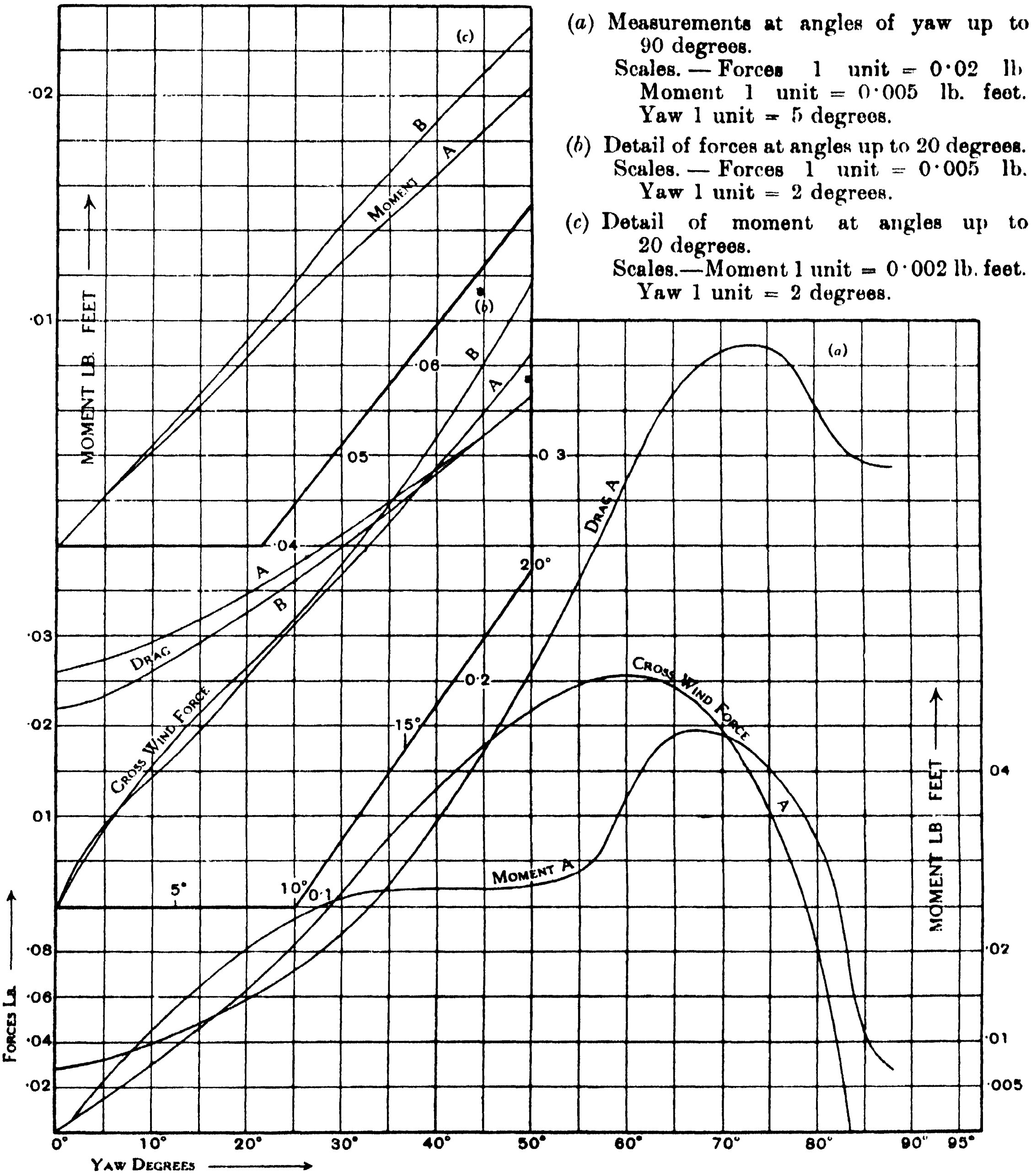
These equations are of the most natural forms to make $f_R$, $f_L$, and $f_M$ of no physical dimensions. The arguments of § 1.01, by which the form of equation (1.001) was justified to some extent, probably apply with equal force in this more general case. The form chosen is suggested by the aerodynamical treatment of the force system on an aeroplane. Since L and M, by symmetry, vanish with $\delta$, the factor $\sin\delta$ is explicitly included in (1.102) and (1.103), in order that the cross wind force and moment coefficients, $f_L$ and $f_M$, may have non-zero limits as $\delta \to 0$. We shall use the symbols $f_R(v/a), f_L(v/a), f_M(v/a)$ for $f_R(v/a, 0)$, $\mathop{\mathrm{Lt}}_{\delta \to 0} f_L(v/a, \delta)$, and $\mathop{\mathrm{Lt}}_{\delta \to 0} f_M(v/a, \delta)$ respectively, and shall omit the explicit mention of the argument $v/a$ when no confusion can arise by so doing.

In view of the evidence mentioned in § 1.01, we may confidently expect that, for all values of $\delta$, all three coefficients will be nearly independent of $v/a$ in the region $0{\cdot}1 \leq v/a \leq 0{\cdot}7$, and shall, when required, assume their absolute independence of $v/a$

Fig. 2. Force components on the 3-inch shells A and B, measured in a wind channel at a wind speed of 40 f.s., plotted against angle of yaw.

Shell of form A.—Moment measured about a point 4·85 inches from base.

" " B.—Moment measured about a point 4·85 inches from base.



(*a*) Measurements at angles of yaw up to 90 degrees.
Scales. — Forces 1 unit = 0·02 lb.
Moment 1 unit = 0·005 lb. feet.
Yaw 1 unit = 5 degrees.

(*b*) Detail of forces at angles up to 20 degrees.
Scales. — Forces 1 unit = 0·005 lb.
Yaw 1 unit = 2 degrees.

(*c*) Detail of moment at angles up to 20 degrees.
Scales.—Moment 1 unit = 0·002 lb. feet.
Yaw 1 unit = 2 degrees.

when $v/a \leq 0 \cdot 7$. With regard to their dependence on $\delta$ we are not here concerned experimentally with $f_R$. We shall assume for the purpose of analysing our experiments, where only a rough value of $f_R$ is required, that $f_R(v/a, \delta)$ is independent of $\delta$ for small values of $\delta$.* For the usual position of the centre of gravity of the shell, $f_M$ at low velocities is remarkably nearly independent of $\delta$ for all values less than 10 degrees, and then diminishes as $\delta$ increases beyond this value. On the other hand, at low velocities, $f_L(v/a, \delta)$ behaves curiously for small values of $\delta$. The wind-channel value of $f_L(v/a)$ is in consequence uncertain. Typical curves showing $f_R$, $f_L$ and $f_M$ as functions of $\delta$ at low velocities are shown in fig. 2. It is the main purpose of the experimental part of this paper to determine $f_L(v/a)$, and $f_M(v/a)$ as functions of $v/a$, when $v/a > 0 \cdot 7$.

1.11. *The Effect of the Angular Motion of the Axis of the Shell.*—In practice the direction of the axis of the shell relative to the direction of motion changes fairly rapidly. By analogy with the treatment of the motion of an aeroplane, we assume, tentatively, that the components of the force system R, L, and M are unaltered by the angular velocity of the axis, but that the effect of the angular motion of the axis of the shell can be represented by the insertion of an additional component, namely, a couple H, called *the yawing moment due to yawing*, which satisfies the equation

$$(1.111) \qquad H = \rho v w r^4 f_H(v/a, \ldots),$$

where $w$ is the resultant angular velocity of the axis of the shell. The form of (1.111) is chosen to make $f_H$ of no physical dimensions and is the only one suitable for the purpose. The couple H is assumed to act in such a way as directly to diminish $w$ (see fig. 3). The yawing moment coefficient $f_H$ may be expected to vary considerably with $v/a$. It may depend appreciably on other arguments such as $wr/v$ and $\delta$. This couple is suggested by, and is analogous to, the more important of the "rotary derivatives" in the theory of the motion of an aeroplane. It appears from considerations of symmetry that no other couple of the "rotary derivative" type need be considered. We shall arrive at rough values of $f_H$ from our experimental results, and to some degree an *a posteriori* justification of our

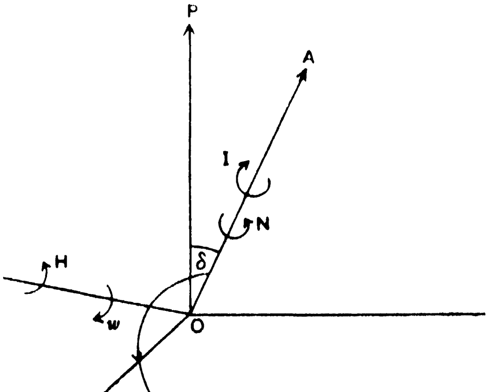


Fig. 3.

---

* By symmetry $\partial f_R/\partial \delta = 0$, when $\delta = 0$, since $f_R$ has a minimum for $\delta = 0$. It might therefore be expected that, when $\delta$ is less than 3 degrees (say), $f_R$ would be nearly independent of $\delta$. This, however, is not the case in wind-channel experiments. The drag at 2 degrees and 3 degrees yaw may be 7 per cent. and 10 per cent. greater, respectively, than the drag at zero yaw. Such evidence as exists indicates that the same increase may occur also at high velocities. An experimental study of the variation of the drag with $\delta$ at high velocities would present no insuperable difficulties with modern apparatus.

assumption that L and M are unaffected by the angular velocity of the axis. But the values we obtain are too rough to enable us to study the variations of $f_H$ with any argument.

1.12. *The Effect of the Axial Spin of the Shell.*—We have so far ignored the possible effect of the spin N of the shell about its axis of symmetry. We shall assume that the preceding components of the force system R, L, M and H are not appreciably affected by this spin. This is in accordance with such evidence as exists in the case of zero yaw (§ 1.01). If, moreover, the component M were seriously affected by the spin, the effect would have been detected by the present trial. No such effect was found (see § 4.13), and this fact provides some evidence of the validity of the above assumption, at least as a first approximation.

The spin N will, however, give rise to certain additional components of the complete force system. There will be a couple I which tends to destroy N, and, when the shell is yawed, a sideways force, which need not act through the centre of gravity, analogous to that producing swerve on a golf or tennis ball. This force must, by symmetry, vanish with the yaw. The swerving force must act normal to the plane of yaw, otherwise it would merely have a component which altered R or L (acting in the plane of yaw), and we have assumed that no such component exists. The complete effects of the spin N can therefore be represented by the addition to the force system of the couples I and J and the force K, acting as shown in fig. 3. To procure the correct dimensions we may assume that those components have the forms*

(1.121) $$\mathrm{I} = \rho v \mathrm{N} r^4 f_{\mathrm{I}},$$

(1.122) $$\mathrm{J} = \rho v \mathrm{N} r^4 \sin\delta\, f_{\mathrm{J}},$$

(1.123) $$\mathrm{K} = \rho v \mathrm{N} r^3 \sin\delta\, f_{\mathrm{K}}.$$

The coefficients $f_{\mathrm{I}}$, $f_{\mathrm{J}}$, $f_{\mathrm{K}}$ may depend effectively on a number of variables which we can make no attempt to specify in the present state of our knowledge. These components may be expected to be very small in comparison with L and M; no certain evidence that they exist is given by our experiments.

1.13. *Relations Between the Components of the Force System.*—The various coefficients in the foregoing specifications will all depend on the external shape of the shell; results obtained for one shape cannot be applied to another. For shells of given shape, however, moving in a given manner, the forces R and L are independent of the position of O, the centre of gravity, while the moment M varies with the position of O. If $\mathrm{M}_1$ and $\mathrm{M}_2$ are the values of M corresponding to positions $\mathrm{O}_1$ and $\mathrm{O}_2$ of O, then

(1.131) $$\mathrm{M}_1 = \mathrm{M}_2 + \mathrm{O}_1\mathrm{O}_2\,(\mathrm{L}\cos\delta + \mathrm{R}\sin\delta),$$

where $\mathrm{O}_1\mathrm{O}_2$ is positive when $\mathrm{O}_1$ is nearer the base than $\mathrm{O}_2$. Using the relations (1.101) to (1.103), and assuming that the yaw is small, the equation (1.131) reduces to

(1.132) $$f_{\mathrm{M}_1} = f_{\mathrm{M}_2} + \frac{\mathrm{O}_1\mathrm{O}_2}{r}(f_{\mathrm{L}} + f_{\mathrm{R}}).$$

This equation is of considerable practical importance, as it enables us to deduce the values of $f_L$ from the values of $f_M$ for two different positions of the centre of gravity. It is found that $f_L$ cannot conveniently be directly observed.

---

* We shall frequently write $\mathrm{I}' = \mathrm{I}/\mathrm{AN}$, where A is the moment of inertia about the axis of symmetry of the shell (see § 1.31).

It will be found convenient in the practical use of (1.132) to introduce the force component *normal to the axis of the shell.* If $f_N$ is the corresponding coefficient, it is easily seen that, when the yaw is small,

$$f_N = f_R + f_L \tag{1.133}$$

and that it is $f_N$ that is directly determined by the variations of $f_M$.

No other relations between the various coefficients are available. Previous to the present experiments, when no definite information existed as to the form of $f_L$ and $f_M$ as functions of $v/a$, special arbitrary assumptions have been made, in order to carry out calculations of the drift of a shell, or of the twisted curve described by its centre of gravity. The authors have made considerable use of the assumptions that the fractions

$$\frac{f_R(v/a,\delta)}{f_R(v/a,0)},\quad \frac{f_L(v/a,\delta)}{f_R(v/a,0)},\quad \frac{f_M(v/a,\delta)}{f_R(v/a,0)}$$

are independent of $v/a$, and have determined their values by wind-channel observations. CRANZ,[*] using essentially the same assumptions, has calculated the values of these fractions by an empirical law due to KUMMER. It must be emphasised that the use of any assumption of this type is of very dubious validity, and that, so far as experiments have yet gone, they have not confirmed any such assumptions. When the values of the coefficients $f_R$, $f_M$ and $f_L$ are required for a shell of any given external shape they can and must be determined by direct experiment.

1.14. In the preceding sections, we have built up, by synthetic arguments, what appears to be the most probable complete force system. It will be seen that in so doing we have actually introduced what can be regarded as a complete system of three forces and three couples referred to three axes at right angles. Owing, however, to the complex nature of the reactions, it appears to us to be essential to construct our force systêm in this manner, instead of attempting to analyse a complete system of three forces and three couples, and assign each component to its proper causes. In this construction, we have been guided by considerations of symmetry, the theory of dimensions, the analogy with the theory of the aeroplane, and also, of course, by the all-important fact that the results of this construction are in agreement with experiments, so far as these have yet been carried. Of our seven components by far the most important are R, L and M; then, some way behind, H. Our experiments were designed to determine L and M, and if possible to throw some light on the size of H, and in these objects a successful start has been made. As a result, it seems reasonable to expect that the preceding specification of the complete force system will prove to be adequate; but much more work on these and other lines[†] is still required. With the numerical knowledge already obtained, which is given in § 1.2, the motion of a shell, of the shape used in these experiments, can be calculated with some approach to certainty. The general nature of the motion is described in § 1.3.

---

[*] 'Zeitschrift für Math. u. Phys.,' vol. XLIII., p. 184.

[†] For instance, the determination of the couple I that destroys the axial spin and the behaviour of $f_R$ as a function of $\delta$.

### § 1.2. *The Numerical Results of the Experiments.*

We now proceed to give the numerical results obtained by analysis of the observations by the methods explained in detail in Part IV.

1.21. *The Values of* $f_{\rm M}$ *and* $f_{\rm L}$.—The observed values of $f_{\rm M}$ and $f_{\rm L}$ are shown plotted against $v/a$, in fig. 4, for the shell of external form A.* The value of $f_{\rm M}$ is

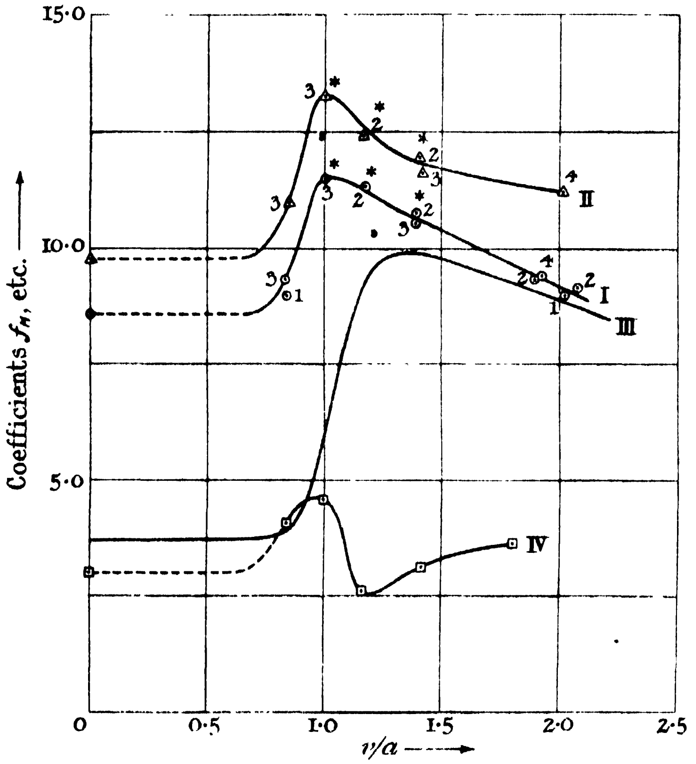


Fig. 4. Shells of form A.

Curve I.—The couple coefficient $f_{\rm M}(v/a)$ for 3-inch shells, with the centre of gravity 4·73 inches from the base.

Curve II.—The same, with centre of gravity 4·20 inches from the base.

Curve III.—The drag coefficient $f_{\rm R}(v/a)$ for comparison on ten times the scale.

Curve IV.—The cross-wind force coefficient $f_{\rm L}(v/a)$.

The plotted points ⊙, △, ⊡ show the observed values. The numbers denote the number of observations whose mean is represented by the plotted point. The stars distinguish those groups fired from the gun rifled one turn in thirty diameters. The others were fired from a gun rifled one turn in forty.

---

* See fig. 6. Form A may be specified thus :—Length 3·84 shell diameters. Base cylindrical. Head with an ogive of 2 diameters radius. Centre of gravity 1·577 diameters from base.

given on the assumption that the centre of gravity is 4·73 inches from the base in the 3-inch shells used. The value of $f_R$ (at zero yaw) is also given for comparison. In fig. 5, the corresponding values of $f_M$ and $f_R$ are given for the shells of external form B,* the centre of gravity being supposed to be 4·965 inches from the base in the 3-inch shells used. These values have been corrected as far as possible for the effect of the cards (see § 2.32), and smooth curves have been drawn through the observations. The values of $f_M$ and $f_L$ for shell A, and $f_M$ for shell B, are given in the following table, Table I., for values of $v/a$ varying by 0·1. These values have been read from the smooth curves of the figures. Besides $f_L$, the value of $f_N$, the force coefficient normal to the shell, is also given. These figures and Table I. represent the main results of the experiment. The values of $f_M$ have a probable error of less than 2 per cent., and the values of $f_L$ of about 10 per cent.

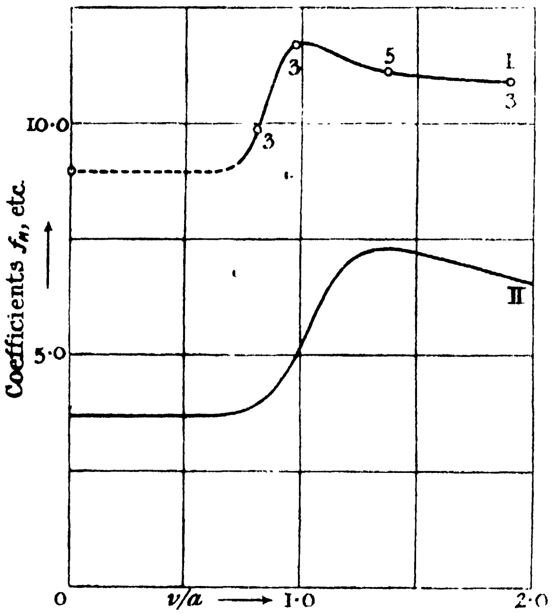


Fig. 5. Shells of external form B.

Curve I.—The moment coefficient $f_M(v/a)$ for 3-inch shells with a centre of gravity 4·965 inches from the base.

Curve II.—The drag coefficient $f_R(v/a)$ shown roughly on ten times the scale.

The differences in the various curves for $f_R$, $f_L$ and $f_M$ are very instructive. They show the complete impossibility of regarding the ratio of $f_R/f_M$, for example, as constant for large variations of $v/a$. Unlike $f_R$, $f_M$ is comparatively unaffected by the velocity of sound. It increases only to about 35 per cent. above its low velocity value, and does not maintain this increase except for a narrow range of velocities near $v/a = 1$. On the other hand $f_R$ increases to two and a-half times its low velocity value and maintains this increase.

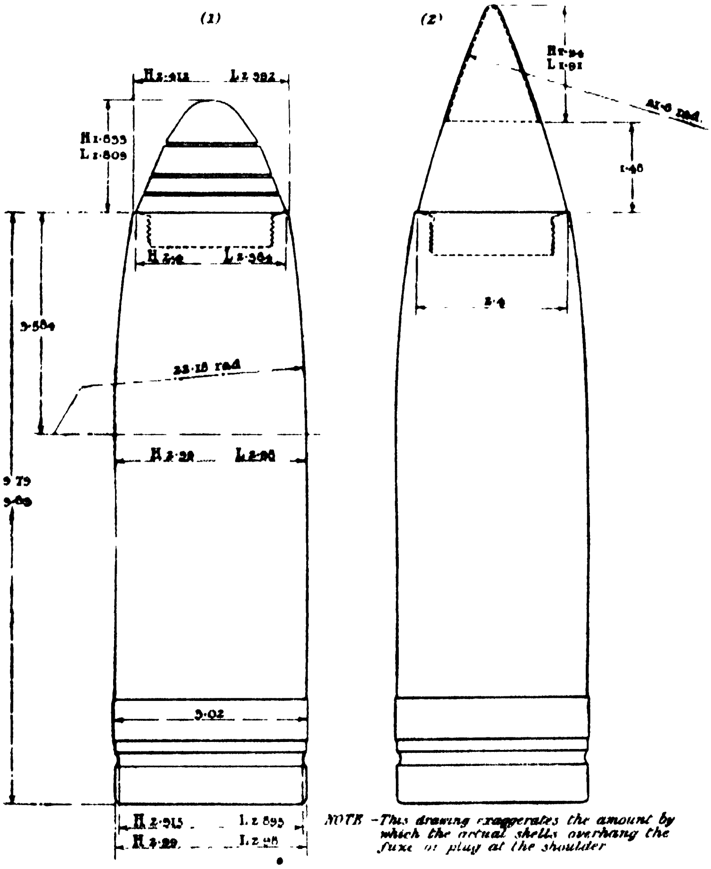
---

* See fig. 6. Form B may be specified thus:—Length 4·34 diameters. Base cylindrical. Head with an ogive of 6 diameters radius. Centre of gravity 1·655 diameters from base.

TABLE I.—Experimental Values of the Couple Coefficient $f_M(v/a)$, the Normal Force and Cross Wind Force Coefficients $f_N(v/a)$ and $f_L(v/a)$, for Shells of Form A, fig. 6; also Values of $f_M(v/a)$ for Shells of Form B, fig. 6.

Determined by firing trials with 3-inch shells.

| $v/a$. | Shell of form A. | | | Form B. |
|---|---|---|---|---|
| | $f_M(v/a)$. | $f_N(v/a)$. | $f_L(v/a)$. | $f_M(v/a)$. |
| Wind channel. | 8·57 | 3·34* | 3·0* | 8·95 |
| 0·7 | 8·6 | — | — | 9·05 |
| 0·8 | 9·05 | 4 3 | 3·9 | 9·75 |
| 0·9 | 10·35 | — | — | 11·15 |
| 1·0 | 11·55 | 5·2 | 4·6 | 11·7 |
| 1·1 | 11·4 | — | — | 11·6 |
| 1·2 | 11·1 | 3·5 | 2·6 | 11·35 |
| 1·3 | 10·8 | — | — | 11·15 |
| 1·4 | 10·55 | 4·1 | 3·1 | 11·05 |
| 1·5 | 10·3 | — | — | 11·0 |
| 1·6 | 10·05 | 4·3 | 3·35 | 11·0 |
| 1·7 | 9·85 | — | — | 10·95 |
| 1·8 | 9·65 | 4·5 | 3·6 | 10·95 |
| 1·9 | 9·4 | — | — | 10·90 |
| 2·0 | 9·15 | — | — | — |



NOTE.—This drawing exaggerates the amount by which the actual shells overhang the fuze or plug at the shoulder

Fig. 6. Showing the external contour of the 3-inch 16-lb. shells, Design H.E. Mark IIB, used in the trial with (1) No. 80 fuze, Mark III; (2) 6 C.R.H. plug, Design 25420.

*Note.*—The driving band is shown cut off at a diameter of 3·02 inches, its mean diameter after engraving.

* Uncertain.

As already mentioned in § 1.1, the low velocity value of $f_L$, as determined in the wind channel, is somewhat doubtful.* There appears to be a distinct minimum in this coefficient soon after the velocity of sound, followed by a steady rise. This type of curve is rather unexpected, and confirmation by a repetition of the experiment is very desirable. We may emphasise again that $f_L/f_R$ is by no means constant, and that $f_L/f_M$ (see fig. 15) undergoes considerable variations.

1.22. *The Force Components* $f_H$ *and Others.*—We have now to exhibit the information obtainable from the damping of the oscillations of the axis. When $f_L$ is known, this information (see § 4.12), provides numerical values for two quantities, one of which is $f_H$ and the other $f_H+\epsilon$, where $\epsilon$ depends on the coefficients $f_J$ and $f_I$ and is *a priori* unlikely to be comparable with $f_H$. The data at our disposal are very rough and could be improved on in future experiments. The present results vary largely in some cases from round to round; the value of $\epsilon$ is much larger than its expected value and of the opposite sign. The general features of the damping (see figs. 12, 14) are however clear and qualitatively consistent. We can assert that the following rough values of $f_H$, given in Table II., are of the right order of magnitude and perhaps not in error by more than 50 per cent. Owing to their roughness they are given for the groups as fired. An attempt has been made to determine $f_H$ in a wind channel at low velocities, the value 22 being obtained.

Table II.—Probable Values of $f_H$, the Coefficient of the Yawing Moment due to Yawing. Groups I., II., III. refer to shells of Form A with various positions of the centre of gravity (see § 2.2). Group IV. refers to Form B.

| Group.<br>Muzzle velocity. | $f_H$. | Group.<br>Muzzle velocity. | $f_H$. | Group.<br>Muzzle velocity. | $f_H$. |
|---|---|---|---|---|---|
| I. 22–24<br>1119 | 80 | II. 24<br>1292 | 70 | III. 1–4<br>2025 | 70 |
| I. 25, 26<br>1326 | 70 | II. 5–7<br>22, 23†<br>1587 | 75 | IV. 13–15<br>1078 | 55 |
| I. 27, 28<br>1563 | 60 | II. 1–4<br>2024 | 60 | IV. 16–18<br>1547 | 75 |
| I. 1–4<br>2167 | 35 | III. 17–19<br>1119 | 40 | IV. 24–26<br>2120 | 80 |
| I. 19–21<br>2320 | 30 | III. 20, 21<br>1292 | 70 | | |
| II. 17–19<br>1119 | 90 | III. 22, 23<br>1567 | 60 | | |

---

* In fig. 4, and subsequently, the low velocity value of $f_L$ is assumed to be $f_L(10°)$ in place of $\underset{\delta\to 0}{\mathrm{Lt}}\, f_L(\delta)$, which is uncertain.

† From guns of different riflings, with results in agreement.

An interesting feature of the damping is that, at a velocity of about 900 f.s., the yaw has a distinct tendency *to increase* (*instead of decreasing*) *with the time; this happens with all four types of shells.* Whether this represents a real phenomenon or is caused by the impacts on the cards (§ 4.5) is not yet clear. It is not physically impossible that $f_H$ may be negative for this velocity. These rounds are ignored here, and further details must be postponed for Part IV.

### § 1.3. *A Description in General Terms of the Angular Motion of the Axis of a Shell.*

The numerical values of $f_R$, $f_L$ and $f_M$ described in § 1.21, with the addition of the rough values of $f_H$ given in § 1.22 make it possible to determine numerically, by the principles of rigid dynamics, the motion of a shell projected in any manner, provided that the velocity ratio $v/a$, and the angle of yaw $\delta$, do not pass outside the limits for which the determination is valid. It is necessary to obtain and solve the dynamical equations of motion in terms of the force components before proceeding to the inverse process of deducing the forces from the observed motion of the shell. Before doing so, however, it is convenient to describe in general terms the motion of the shell in various circumstances; this description is qualitative only, and is inserted for the purpose of illustration: the quantitative results are reserved for Part IV.

1.31. *The Spinning Top Analogy.*—We have already noticed in the Introduction the important analogy between the motion of the axis of a shell and the axis of a spinning top. With the reservations there made, the analogy is complete, so long as $f_M$ can be regarded as independent of $\delta$. The equations of motion of a stable shell, given in § 3.2, are a generalisation of the equations for the small oscillations of a top in the neighbourhood of the vertical. For the general case of stable or unstable motion where the yaw need not be small, some use can be made of the exact equations of motion of the top (§ 3.4).

In particular, the condition for *the stability* of a shell is identical with the condition for a top. The condition that the shell should be in stable equilibrium with its axis parallel to its direction of motion is that

$$A^2N^2 > 4B\mu, \tag{1.311}$$

where A and B are, respectively, the moments of inertia of the shell about longitudinal and transverse axes through the centre of gravity, N is the spin of the shell about its (longitudinal) axis in radians per second, and $\mu \sin \delta$ is equal to M, the moment of the air forces about the centre of gravity. It is therefore convenient to define a new variable $s$, "the coefficient of stability," by the equation

$$s = A^2N^2/4B\mu. \tag{1.312}$$

When $s$ is greater than unity by a sufficiently large amount, a possible form of angular motion for both shell and top consists of a small oscillation, composed of periodic terms with two distinct periods. The values of these two periods are uniquely determined by the values of $s$ and $AN/B$ or $\Omega$; conversely $s$ and $\Omega$, and hence $\mu$ and $f_M$ are uniquely determined by the values of the periods. The main object of the jump card experiment, described in this paper, is to determine the two periods of the initial angular oscillations of a shell, fired horizontally from a gun. As $\Omega$ depends only on the spin N (known in terms of the muzzle velocity) and the moments of inertia, there is in general an independent check on the observation.* By firing the shell at a series of different muzzle velocities, values of $f_M$ are determined for different values of the variable $v/a$, resulting in the curves of § 1.21.

1.32. The success of the experiments depends entirely on the occurrence of accidental disturbances at the muzzle, in order to produce oscillations of sufficient amplitude to be measurable. The methods of observation used were capable of giving accurate results, provided that the maximum yaw exceeded 1 degree. In the actual trial, no round was fired which developed a maximum yaw of less than 2 degrees, and it is probable that with almost any type of shell the initial disturbance would be sufficient for observations of this nature to be made. It may be noticed that, for a given initial disturbance, the amplitude of the oscillations is greater, the smaller the value of $s$, until, as $s$ approaches and becomes smaller than the value unity, the amplitude of the oscillations increases very rapidly. For this reason it was at first considered preferable to deal with a shell and gun for which $s$ was only just greater than unity, but the experiments described in this paper indicate that a value of $s$ in the neighbourhood of $1 \cdot 5$ will give the best general results.

It is to be expected *a priori*, and is confirmed by the experiment, that the initial yaw of a shell, on leaving the muzzle of a gun, is very small, and that the angular oscillations are due mainly to an initial angular velocity about a transverse axis. The shell is completely unstable under the very large pressures of the powder gases on its base, so that as soon as it is released from the barrel it is disturbed from its position of unstable equilibrium by an amount, and in a direction, which depend largely on accidental circumstances.† The pressure of the powder gases probably continue to influence the motion over a short interval after the shell has left the gun, but the whole effect on the shell must approximate to that of an impulsive couple about a transverse axis.

The angular motion of the shell, for some distance from the muzzle, approximates, therefore, to the type of motion of a spinning top known as rosette motion, in which the axis of the top passes periodically through the vertical.

---

* This check is especially important in the case of shells of type II., as the shift of the lead block on firing alters the values of the dynamical constants as determined by laboratory experiments (§ 2.2).

† [*Note added July* 31, 1920. In view of further analysis of the initial circumstances of shells in this trial, this account of the matter is probably incomplete.]

1.33. *Differences Between the Shell and Top Movements.*—We now proceed to consider the factors, so far neglected, which cause the angular motion of the shell to differ from that of the corresponding top. These may be enumerated as follows:—

(1) The effect of the cross-wind force in causing the centre of gravity to follow a curve of helical type.
(2) The effect of the force components denoted in § 1.11 and § 1.12 by H, J, and K.
(3) The effect of the diminution of forward velocity caused by the drag.
(4) The effect of gravity.

These effects will be considered in turn.

1.331. The angular oscillations of the shell give rise to a cross-wind force, which varies in magnitude and direction as the yaw varies, and this modifies the straight line motion along the direction of projection into motion of a helical type. If this helical motion could be observed with accuracy it would give valuable data for the cross-wind force coefficient $f_L$, but unfortunately the amplitude of the oscillations is too small to allow of this. Hence the most important effect, from the point of view of these experiments, is the reaction of the sideways motion of the centre of gravity on the angular oscillations of the shell. This helps to damp out the oscillations.

1.332. The yawing moment factor H has a similar damping effect as it is always opposed to the transverse angular velocity. While the effect of the former factor is to damp the slow period oscillation and slightly augment the quick oscillation, this latter has exactly the reverse effect. In combination, they, in general, damp out the oscillations of both periods. For the 3-inch shells, used in this trial, the yawing moment damping factor is of greater importance than the cross-wind force damping factor, and the general effect is to diminish the maximum values of the yaw, and at the same time to convert the initial rosette motion into the slower steady precessional motion.* The force component, J, due to the spin, has no appreciable effect on the angular motion, but the corresponding couple K might act as a small additional damping factor.

1.333. The head resistance or drag slowly diminishes the forward velocity, and so increases the stability factor $s$, by diminishing $\mu$. The change in $s$ diminishes the amplitude of the oscillations to a limited extent, and so assists the other damping factors.

1.334. Gravity affects the angular motion of the axis of the shell by producing curvature in the trajectory. In taking account of the gravity effect it is necessary to refer the motion to axes moving with the tangent to the trajectory (see § 3.2). The effect is quite insignificant over the range covered by the present trial, but becomes of importance at later stages of the trajectory, where it is responsible for producing the drift.*

---

* There are two possible types of steady precessional motion at constant yaw, one with a quick and the other with a slow precessional velocity.

It is convenient to illustrate this effect by considering a simple case of steady motion.

1.34. *An Illustration of the Gravity Effect.*—Let the centre of gravity of a shell be constrained to move through air at a constant speed $v$, in a vertical circle (fig. 7), the inclination of the path to the horizontal being $\theta$ at any instant. Thus $v$ and $d\theta/dt$ are constant. There is a possible steady motion in which the axis OA always lies in the plane through OP perpendicular to the plane of the circle, the angle AOP ($\delta$) being constant. The couple M tending to increase $\delta$ will also be constant, so that the contemplated motion is the same as the steady motion of a top making an angle $\frac{1}{2}\pi+\delta$ with the vertical which corresponds to the normal to the plane of the circle. The angular velocity of the axis about this normal is $-\theta'$; the value of $\delta$ as given by the ordinary formula for the steady motion of a top under these conditions† is

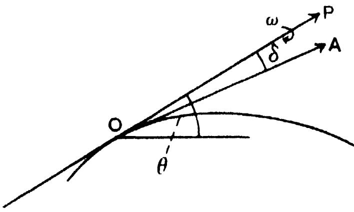


Fig. 7.

$$(1.341)\qquad -\mathrm{AN}\theta'\cos\delta+\mathrm{B}\theta'^2\sin\delta\cos\delta = \mathrm{M} = \mu\sin\delta.$$

If $\theta'$ is not too large and $\mu$ is not too small, a possible value of $\delta$ is small; we may now regard $\mu$ as independent of $\delta$, and the equation then reduces to

$$(1.342)\qquad \delta = -\mathrm{AN}\theta'/\mu = -4s\theta'/\Omega,$$

the term neglected being of order $\delta^3$. When a shell is moving freely the angular velocity $\theta'$ increases, and the linear velocity diminishes up to a point beyond the vertex of the trajectory. If the initial motion is identical with the above steady motion, this will cause the couple M to diminish, so that the axis of the shell will lag behind its position in the steady motion. This lag gives rise to a component angular velocity of the axis tending to increase the yaw $\delta$, until a state of relative equilibrium is reached, in which the yaw is slightly less than its equilibrium value, and the axis lags slightly behind (*i.e.*, above) the tangent OP. When the velocity is high and the spin N not too large, M is large and the true position of the axis lies very near the equilibrium position. It will be shown in fact, in Part IV., that the assumption that OA is level with OP, and that $\delta$ is given by (1.342), lead to a determination of the drift which is sufficiently accurate for all trajectories of elevation less than 30 degrees. The drift is produced by the cross-wind force resulting from the above value of the yaw.

---

* For a shell whose spin and direction of motion are related like a right-handed screw the drift is to the right of the plane of fire.

† See, *e.g.*, ROUTH, 'Rigid Dynamics,' vol. II., Art. 207.

In the neighbourhood of the vertex of a trajectory of high elevation, both the velocity and the couple M become very small, so that $\delta$ becomes large. A calculation has been made, by a step-by-step process, of the angular motion and drift of a shell fired at an elevation of 70 degrees. The yaw, soon after the vertex, reaches the value 60 degrees, while the axis lags behind the tangent to the path by more than 45 degrees.

1.35. The effect of gravity as described in the last section completes the list of factors which have an appreciable effect on the motion, and it remains to consider the way in which they combine. It will be shown in § 3.2 that the motion of the axis of a stable shell is determined, to a good approximation, when the yaw is not too large, by a linear differential equation of the second order. The effect of gravity is to produce the type of motion described in § 1.34, given the proper initial conditions in which the yaw and its rate of increase are both very small. The complete motion under arbitrary initial conditions may be obtained by superposing the appropriate type of initial oscillatory motion, which is unaffected by gravity. The superposed oscillations will ultimately be damped out, leaving the motion of the last section only. The motion of the centre of gravity will be appreciably affected by alteration of the initial conditions only in so far as they produce a certain small sideways displacement and velocity (§ 4.2), and increase the drag to an extent which is not yet known.

More detailed results are reserved for Part IV., following the discussion of the mathematical theory. Actual examples of the observed motion of the shell's axis can be studied in fig. 14.

## PART II.—DETAILS OF THE EXPERIMENTAL ARRANGEMENTS AND MATERIAL.

### § 2.0. *General Arrangements.*

We propose, in this part, to explain the details of the experiments in so far as is necessary to enable the reader to understand the method used, and to form an estimate of the accuracy obtained, or capable of being obtained, in this manner.

The experiments were carried out as the weather served in January and February, 1919, four different types of 3-inch shells being fired, at various velocities, from each of two differently rifled guns. The constants of the shells used are given in Table III.,

TABLE III.—Mean Values of the Dynamical Constants of the Shells used, Determined before Firing.

Types I., II. and III., Form A. } (See fig. 6.)
Type IV., Form B.

| Type of shell. | Length, inches. | Weight, lb. | Distance of centre of gravity from base, inches. | Axial moment of inertia, A, lb. (in.)$^2$. | Transverse moment of inertia, B, lb. (in.)$^2$. | B/A. |
|---|---|---|---|---|---|---|
| I. (Normal) . . | 11·53 | 14·09 | 4·727 | 18·37 | 143·9 | 7·83(5) |
| II. (Centre of gravity forward) | 11·53 | 16·31 | 5·124 | 19·20 | 165·0 | 8·59 |
| III. (Centre of gravity back) | 11·53 | 16·48 | 4·203 | 18·93 | 129·5 | 6·84 |
| IV. (Shells with pointed nose) | 13·15 | 14·62 | 4·965 | 18·71 | 166·2 | 8·89 |

and details of the groups fired are given in Table IV. The distance available between the firing point and the sea at Portsmouth is rather less than 600 feet. The motion of the shell was recorded over this range, within which the effects of gravity are fairly small and the path of the shell not widely different from a straight line. To achieve this the shell was fired through a series of millboard pistol targets, 2 feet square, about $\frac{1}{20}$ inch thick, which were fastened approximately at right angles to the path of the shell, at suitable distances from the muzzle.* The plane of the card was carefully adjusted, and it is probable that in no case did the angle between the path of the shell and the plane of the card differ from a right angle by as much as two degrees. As errors up to four degrees do not affect the shape and position of the hole in the card, which determine the position of the axis and the centre of gravity of the shell at the moment of impact, it may be assumed that in every case

---

* For the gun whose rifling made one complete turn in a length of 40 diameters of the bore (rifled 1 in 40) ten cards were used, placed approximately at 60-foot intervals, the first card being 50 feet from the muzzle. For the gun rifled 1 in 30 twelve cards were used, the first seven being at 30-foot intervals and the later cards at 60-foot intervals as before. The distance of the cards from the muzzle of the gun was determined with a probable error of 1 inch.

## TABLE IV.—Showing Groups of Rounds Fired.

The types of shell are numbered I.–IV., and the shells of each type are numbered 1, 2, 3, ... in the order of firing.*

GUN rifled one turn in 40 diameters of the bore.

| Group. | Mean muzzle velocity for group, f.s. | Remarks. | Group. | Mean muzzle velocity for group, f.s. | Remarks. |
|---|---|---|---|---|---|
| Type I.; shells of form A; centre of gravity normal. | | | | | |
| I. 11–14 | 922 | Stable | I. 15, 16 | 2130 | Stable † |
| I. 8–10 | 1072 | Unstable | I. 1–4 | 2167 | Stable |
| I. 17, 18 | 1312 | Unstable | I. 19 | 2272 | Stable |
| I. 5– 7 | 1565 | Just stable | I. 20, 21 | 2346 | Stable |
| Type II.; shells of form A; centre of gravity forward. | | | | | |
| II. 8–10 | 934 | Stable | II. 5–7 | 1585 | Stable |
| II. 11–13 | 1107 | Unstable | II. 1–4 | 2024 | Stable |
| II. 14–16 | 1334 | Unstable | | | |
| Type III.; shells of form A; centre of gravity back. | | | | | |
| III. 8–10 | 931 | Stable | III. 5–7 | 1583 | Just stable |
| III. 11–13 | 1077 | Unstable | III. 1–4 | 2025 | Stable |
| III. 14–16 | 1312 | Unstable | | | |
| Type IV.; shells of form B; centre of gravity normal. | | | | | |
| IV. 10–12 | 884 | Very unstable | IV. 1–6 | 2130 | Very unstable |
| IV. 7– 9 | 1553 | Very unstable | | | |

---

* Only the stable groups are analysed in this report. For a specimen yaw curve in an unstable case, see fig. 12.

† Fired with cards on the far screens only, to determine by comparison the effect of the impacts on the cards.

TABLE IV. (continued).

GUN rifled one turn in 30 diameters of the bore.

All groups were stable.

| Group. | Muzzle velocity, f.s. | Group. | Muzzle velocity, f.s. | Group. | Muzzle velocity, f.s. |
|---|---|---|---|---|---|
| Type I. | | | | | |
| I. 22–24 | 1119 | I. 25, 26 | 1326 | I. 27, 28 | 1563 |
| Type II. | | | | | |
| II. 17–19 | 1119 | II. 24 | 1292 | II. 22, 23 | 1589 |
| Type III. | | | | | |
| III. 17–19 | 1119 | III. 20, 21 | 1292 | III. 22, 23 | 1567 |
| Type IV. | | | | | |
| IV. 21–23<br>IV. 13–15 | 900<br>1078 | IV. 16–18<br>IV. 19, 20* | 1547<br>1547 | IV. 24–26 | 2121 |

the centre of gravity of the shell was moving normally to the card.† Thus the angle actually recorded by the shape of the hole in the card is the true yaw of the

---

* Fired with cards on the far screens only, to determine by comparison the effect of the impacts on the cards.

† The angular motion of the axis of the shell is comparatively so slow that it can be ignored during the interval in which the shell is passing through a card. For instance, with the shells used in this trial the change in $\phi$, the orientation of the yaw, is never as much as $3\frac{1}{2}$ degrees during the complete passage through the card, and the change in $\delta$ never as much as 8 minutes. These quantities are of the same order as the errors of observation and may be ignored. Thus the shell can correctly be regarded as equivalent for cutting purposes to its circumscribing cylinder (of indefinite length) whose generators are parallel to the direction of motion of the centre of gravity.

If the direction of motion is normal to the plane of the card at the moment of impact, a certain hole will be cut in the card, whose shape will be precisely that of the normal cross-section of this circumscribing cylinder. But if the card is tilted through a small angle $\tau$ about any axis in its own plane, the hole made by the shell will be the same as the cross-section of the supposed cylinder by the plane of the card shell, that is, the angle between the axis of the shell and the direction of motion of its centre of gravity.

On each card there was marked, by methods which need not be particularised, (*a*) the vertical, (*b*) a reference point from which the point of aim for each round could be deduced. The probable error in the marking of the vertical was negligible compared to the other errors of observation. The probable error* in each co-ordinate of the point of aim was about 0·2 inches.

Times of flight from the muzzle to each card were not directly observed, but the mean velocity of the shell over a suitable interval of the range was observed for each round with two standard Boulangé chronographs. These were sometimes used as a pair—in these cases their readings were in good agreement—and sometimes separately, at opposite ends of the range, to determine the loss of velocity, and so an approximate value for the average coefficient of the drag. From the data so obtained, the muzzle velocity and the times of flight from the muzzle to each screen were calculated by the usual ballistic methods to a nominal accuracy of 1 f.s. and $10^{-4}$ second, respectively. It is improbable that any of these quantities are appreciably in error to the order of accuracy required by the rest of the experiment. A check on the calculated muzzle velocity is provided by the observations, for a discussion of which the reader should refer to § 4.1.

### § 2.1. *Measurement of the Holes in the Cards.*

It is now necessary to deduce, from the position and shape of a hole in any card, the position of the axis and centre of gravity of the shell at the moment of passing the card. This can usually be done with considerable accuracy. It has been found that at all velocities less than 1600 f.s., and often at higher velocities, the hole has the form shown diagrammatically in fig. 8, and by photographs of actual examples in fig. 8A.

Inside the outer circumference ABA′B′ of the hole, a considerable amount of bruised and partly torn card QQQ is left, which is still attached to the untouched part. It is found that, when the edges of this part are flattened out, they always define with some accuracy a circle of diameter 2·40 inches. A stiff paper circle of this diameter can be fitted to the hole with such certainty that its centre is seldom in doubt by more than 0·01 or at most 0·02 inches.

in its tilted position. The dimensions of such a hole will only differ from those of normal impact by terms of order $d(1-\cos\tau)$, where $d$ is any dimension of the hole. Such second-order terms are completely negligible if $\tau < 4$ degrees. Thus in all cases the shell may be regarded as cutting the hole in the card as if the direction of motion of its centre of gravity is normal to the plane of the card at the moment of impact.

* Throughout this paper "probable error" is used with its technical meaning, see *e.g.*, BRUNT, 'The Combination of Observations,' p. 30.

The external form of the shells used in the trial is shown in fig. 6. It will be observed that at the junction of the body of the shell and the fuze or plug there is a distinct cutting edge of plan diameter 2·402 inches. It is clear, therefore, that when the impact takes place, a circle of cardboard, 2·40 inches in diameter, is punched out and cleanly removed by this edge; the greater part of the circumference of this inner circle is usually removed by the subsequent passage of the body of the shell, which cuts the complete hole, but enough remains, in a bruised state, for yaws that are not too large, to define the position of the centre of this section of the shell at the moment of impact on the card.

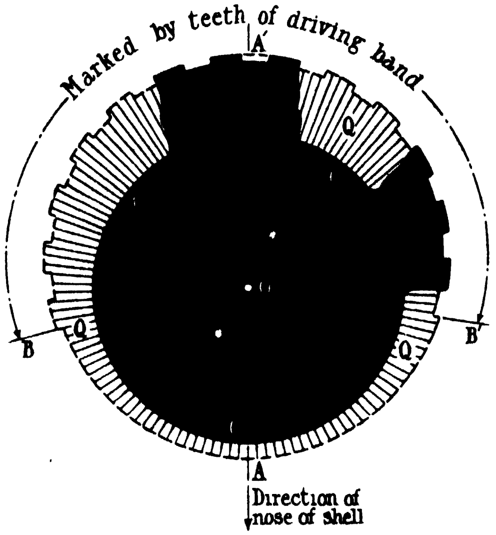


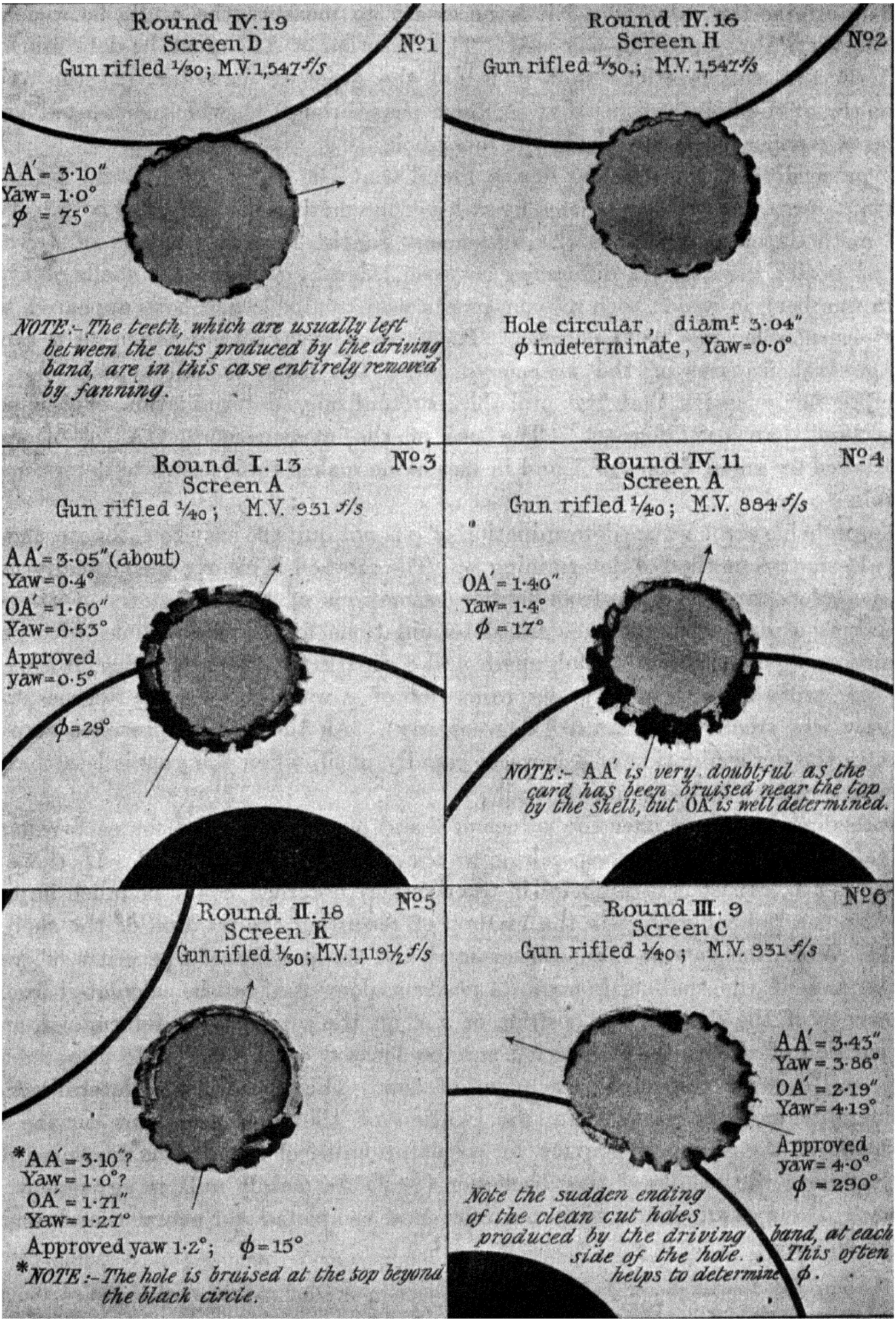
Fig. 8. Diagrammatic sketch of a typical hole, for a yaw between 1 degree and 4 degrees, when the velocity is low or medium.

| | |
|---|---|
| CCC. | Inner circle—radius 2·40 inches, centre O. |
| ABA′B′. | Outer circumference of hole. |
| QQQ. | Bruised part of card. |
| AA′. | Axis of symmetry or greatest diameter of hole. |
| BA′B′. | Circumference cut by teeth of driving band. |
| BAB′. | Ditto cut by nose or shoulder of shell. |

The lengths AA′ (3·16 inches in figure) and OA′ (1·80 inches) each serve to determine the size of the yaw.

The values of the yaw corresponding to the above values are 1·6 degrees and 1·8 degrees respectively, mean 1·7 degrees.

It follows, therefore, that there are two distinct methods by which the value of the yaw $\delta$ can be determined. In the first place, there is a unique relation between the greatest diameter of the hole (AA′, fig. 8) and the value of $\delta$; secondly, there is a unique relation between OA′ and $\delta$. These relations can be tabulated numerically when the plan dimensions of the shell are known, and the value of $\delta$ corresponding to any measured length AA′ or OA′ read off.



Round IV. 19
Screen D
Gun rifled 1/30; M.V. 1,547 f/s

*NOTE:—The teeth, which are usually left between the cuts produced by the driving band, are in this case entirely removed by fanning.*

Round IV. 16
Screen H
Gun rifled 1/30; M.V. 1,547 f/s

Hole circular, diam$^{r}$ 3·04″
$\phi$ indeterminate, Yaw = 0·0°

Round I. 13
Screen A
Gun rifled 1/40; M.V. 931 f/s

Round IV. 11
Screen A
Gun rifled 1/40; M.V. 884 f/s

*NOTE:—AA′ is very doubtful as the card has been bruised near the top by the shell, but OA′ is well determined.*

Round II. 18
Screen K
Gun rifled 1/30; M.V. 1,119½ f/s

**NOTE:—The hole is bruised at the top beyond the black circle.*

Round III. 9
Screen C
Gun rifled 1/40; M.V. 931 f/s

*Note the sudden ending of the clean cut holes produced by the driving band, at each side of the hole. This often helps to determine $\phi$.*

Fig. 8A.

To determine the value of $\phi$,* it is necessary to measure the angle between AA and the vertical recorded on the card. The direction of AA′ must be determined by eye from the considerations that it is (1) the greatest diameter, (2) the axis of symmetry of the hole, and (3) that it must pass through O, which is located as the centre of a paper circle fitted into the inner hole.

By proceeding in this manner it was found that the values of $\delta$ could be nearly always determined with confidence by at least one method and often by both. When both methods were available the agreement in the resulting values of $\delta$ was in general good; the average difference between them in *all* cases for shells of type I. (99 in number), in which both measurements were available and both appeared to be *a priori* reliable, was 0·20 degrees. These cases were simply taken as a sample. The general features of the agreement were the same for all types. We may therefore fairly assert that the probable error of any determination of $\delta$ is something less than 0·2 degrees. The use of the measurement OA′ is of special importance for small values of $\delta$, and in fact alone makes their accurate determination possible.

The probable error in the determination of $\phi$ is not quite so easy to estimate, as there is no alternative method of determining $\phi$. The method is clearly theoretically sound, and the errors can only arise from faulty estimations of the symmetry of the hole. By making a number of independent determinations for the same hole, with proper precautions against a biassed judgment, and comparing their consistency, it appeared that the probable error of any determination of $\phi$ was less than 2½ degrees, unless the yaw was small (less than 0·8 degrees, say). As the yaw approaches zero, the errors in the determination of $\phi$ increase rapidly until, when the yaw is less than 0·2 degrees, $\phi$ cannot be determined at all.

Proceeding in this manner the values of $\delta$ and $\phi$ were tabulated for each round for the values of the time corresponding to the position of each card. If the above estimates are correct it is doubtful if the accuracy obtained could be much improved on without a radical change in the method of recording the position of the shell.

2.11. When the yaw has been determined, and the position of the centre of gravity on the axis of the shell is known, its position along AA′ can be calculated from the dimensions of the shell. The position of AA′ on the card is well determined, and so the position of the centre of gravity can be located with respect to the reference point, and so with respect to the point of aim. This part of the determination is considerably more accurate than the location of the reference point on the card. The path of the centre of gravity for a small number of rounds was measured up in this manner; the results of the discussion (§ 4.2) are mainly null, in agreement with theory. The measurements were therefore not completed for every round and are not given here.

---

* The angle $\phi$ denotes the angle between the plane of yaw OAP and the vertical plane through OP. See fig. 10, p. 332.

### § 2.2. *Determination of the Dynamical Constants.*

All the shells used in these experiments were weighed before firing, and their overall lengths were measured. The variations from shell to shell were small, and the mean values given in the tables may be assumed to be correct for all purposes. No appreciable change in these quantities is likely to occur on firing.

The moments of inertia were determined, before firing, for a selection of about 25 per cent. of the shells of each type. The probable error of any determination was about 1 part in 2000. The mean values for the different types of shell are given in the table. The extreme variation of any transverse moment of inertia from the mean was 1·8 per cent., and of the axial moment of inertia was 0·8 per cent. The errors in assuming that the mean value of the sample is the correct value for each round may therefore be appreciable at times, but should not seriously affect the final mean results. The general accuracy of the experiments was, contrary to expectation, sufficient to warrant the refinement of determining and using the individual values for each shell.

The centres of gravity were also determined, before firing, for the same selection of shells, and the mean value of the distance of the centre of gravity from the base is given, in the same table, for each type. The determination was made with a probable error of 0·003 inches. The values were fairly constant for the shells of any one type, the extreme variation from the mean being 0·022 inches.

It is by no means certain *a priori* that the values of A and B and the position of the centre of gravity may not be changed appreciably in some of the shells by the stresses set up when the gun is fired. No change is at all likely in the empty shells of types I. and IV., or in the bodies of the other shells; they may be confidently relied upon not to be stressed beyond their elastic limit; but the lead and wood filling in the shells of types II. and III. is decidedly suspect. To test this point, two shells of each of the types II. and III., after the determination of their dynamical constants, were fired* over water for recovery, and their constants were then re-determined. In the case of the shells of type III., with a filling of lead at the back and wood in front, there was no appreciable change. In the case of the shells of type II. with lead in front and wood behind, the wood block, as might have been expected, was crushed, and the lead had moved back about three inches in the case of the high velocity and one inch in the case of the low. The axial moments of inertia, A, were unaltered, but the transverse moments of inertia B and the positions of the centre of gravity were of course seriously affected. It was found, however, that the observed changes in both could be satisfactorily accounted for by the observed movement of the lead block, of weight 1·9 lb. When the centre of gravity of the shell of type II. is 4·727 inches from the base, so that it coincides with the centre of gravity of a shell of type I., the value of B is 145·7 lb. (inch)$^2$. Neglecting the effect of the wood, suppose that the lead plug is $x$ inches further forward. In such a case

---

* One of each type at a muzzle velocity of 1950 f.s. and one at 1530 f.s.

$$B = 145{\cdot}7 + 1{\cdot}9x^2. \tag{2.21}$$

If, moreover, $l$ is the distance of the centre of gravity from the base in inches, then

$$l = 4{\cdot}727 + 0{\cdot}117x. \tag{2.22}$$

The altered position of the centre of gravity can therefore be recovered by calculation, if the altered value of B can be deduced from the observations. This is, in fact, the case (see § 4.1), so that even for shells of type II. the dynamical constants of the shells after firing are satisfactorily certain.

### § 2.3. *Possible Disturbing Factors.*

There are two further possible causes of error which we have not yet mentioned. These are (1) the wind, and (2) the impulsive action between the shell and the card.

2.31. *The Effect of Wind.*—Since we are studying the motion of the shell under the force system impressed by the air, we are concerned solely with the motion of the projectile *relative to the air*, but we can only observe, by means of jump cards, the motion of the projectile *relative to the ground.*

If the strength and the direction of the wind are known, it is an easy matter to convert the observed values of the size and orientation of the yaw, and the observed motion of the centre of gravity, into the corresponding quantities for the motion relative to the air. It is, however, very difficult to determine what is the strength of the wind, at the moment of firing, only a few feet above the ground. It is, therefore, necessary to carry out jump card trials in calm weather. During the experiments the wind exceeded 10 f.s., only at the moments of firing three rounds, and was usually only 5 or 6 f.s. at 20 feet above the ground. Its strength near the ground will have been still less, and its effects may therefore be neglected.

2.32. *The Impulsive Action between the Shell and the Card.*—When the experiments were started it was not expected that the effect of the cards would be decidedly bigger than the probable random errors of the results. This, however, appears to be the case. A limited amount of evidence, for determining the necessary correction, is supplied by the few comparative rounds fired without cards on the nearer screens. Such comparative rounds would have been included in all, or at least the majority, of the groups, if their importance had been realised earlier. The evidence supplied by the comparative rounds was carefully analysed, and was supplemented, after the conclusion of the trial, by determination of the *magnitude* of the impulse between the cards and the shells by observation of the extra loss of velocity so caused. The magnitude of a single impulse, at not too great a value of the yaw, probably has the values

14·3 foot-poundals at 2470 f.s.,

8·9 foot-poundals at 1140 f.s. ;

the values at other velocities may be roughly obtained by linear interpolation.

The effect on the observed motion of the axis due to an impulsive couple was calculated, and it was found that rough values could be assigned for the magnitude of the impulsive couple acting at any card. On calculating the total effect on the observed value of $s$ it was found that the probable correction required varied from $2\frac{1}{2}$ to $4\frac{1}{2}$ per cent. in the various groups. This correction was applied before constructing Table I. and figs. 4 and 5. The figures of Table II. have not been corrected for this effect as their accuracy is not great enough to make it worth while to do so.

## PART III.—METHODS OF OBTAINING AND SOLVING THE EQUATIONS OF MOTION OF A SPINNING SHELL.

### § 3.0. *Introductory.*

On the assumptions discussed in Part I. the equations of motion of a spinning shell can be written down at once by the rules of rigid dynamics. Three different types of these equations will be found of use in practice, all of which may be obtained most simply as special cases of the vector equations of motion of the shell, referred to axes rotating in the most general manner. The use of the vector notation, in the initial stages of the discussion, has the further advantage of showing most clearly the meaning of the various terms, and of presenting the results in a symmetrical form.

In order to simplify the general equations, the only components of the force system impressed by the air, retained in the initial discussion, are R, L, M, and the spin-retarding couple I ($= \mathrm{AN}\Gamma$). The remaining components are of less importance and will be inserted later on in § 3.5.

After obtaining the general equations the three special types are deduced. They may be described as follows :—

*Type* $\alpha$.—Equations in terms of direction cosines, referred to axes moving with the tangent to the corresponding plane trajectory.

*Type* $\beta$.—Equations in terms of direction cosines or spherical polar co-ordinates, referred to axes moving with the tangent to the actual twisted trajectory.

*Type* $\gamma$.—Equations similar to the equations of energy and angular momentum of a top (spherical polar co-ordinates), referred to the axes used for type $\beta$.

In each case the equations obtained are simplified by certain approximations, and the results are suitable for use only under certain conditions. Equations of type $\alpha$ are valid when the shell is sufficiently stable and the yaw is small; type $\beta$ when the shell has settled down to a non-periodic motion in which the yaw may be large, the initial oscillations being damped out; and type $\gamma$ when the motion of the centre of gravity is nearly rectilinear.

Equations of these types cannot be solved exactly, and the method of approximation used to obtain a solution is different in each case. The equations of type $\alpha$ are used for the analysis of the jump card experiments, for all sufficiently stable rounds, and could be used to compute the entire motion in any trajectory whose initial elevation is less than 45 degrees. Equations of type $\beta$ have been used to compute the latter part of a twisted trajectory at an elevation of 70 degrees. Equations of type $\gamma$ have a limited application in analysing the jump card records for rounds which are nearly or quite unstable.

3.01. *Note on the Vector Notation.*—All letters which represent vector quantities will be in clarendon type, to distinguish them from scalar quantities in the ordinary type. The three components of any vector $\mathbf{A}$, referred to right-handed rectangular axes 1, 2, 3, are written $A_1$, $A_2$, $A_3$.

If $\mathbf{A}$ and $\mathbf{B}$ are two vectors, their vector product is denoted by $[\mathbf{A}\,.\,\mathbf{B}]$. This represents the vector whose components are

$$(A_2B_3-A_3B_2), \qquad (A_3B_1-A_1B_3), \qquad (A_1B_2-A_2B_1).$$

It is perpendicular to the plane containing the two vectors in the direction of the axis of the right-handed screw, which turns from $\mathbf{A}$ to $\mathbf{B}$, its modulus being equal to the product of the moduli of $\mathbf{A}$ and $\mathbf{B}$ into the sine of the angle between them. The scalar product of the two vectors is written $(\mathbf{A}\,.\,\mathbf{B})$, and is equal to the scalar quantity

$$A_1B_1+A_2B_2+A_3B_3;$$

it is also equal to the product of the moduli of $\mathbf{A}$ and $\mathbf{B}$ into the cosine of the angle between them, being positive when this angle is acute. For simplicity, we denote $(\mathbf{A}\,.\,\mathbf{A})$ by $(\mathbf{A})^2$, which is equal to the square of the modulus of $\mathbf{A}$.

Constant use is made of the following identities:—

$$[\mathbf{A}\,.\,\mathbf{A}] = 0,\ ([\mathbf{A}\,.\,\mathbf{B}]\,.\,\mathbf{A}) = 0. \tag{3.011}$$

$$[[\mathbf{A}\,.\,\mathbf{B}]\,.\,\mathbf{C}] = (\mathbf{A}\,.\,\mathbf{C})\,\mathbf{B}-(\mathbf{B}\,.\,\mathbf{C})\,\mathbf{A}. \tag{3.012}$$

$$([\mathbf{A}\,.\,\mathbf{B}]\,.\,[\mathbf{B}\,.\,\mathbf{C}]) = (\mathbf{A}\,.\,\mathbf{B})\,(\mathbf{B}\,.\,\mathbf{C})-(\mathbf{B})^2\,(\mathbf{A}\,.\,\mathbf{C}). \tag{3.013}$$

## § 3.1. *The General Vector Equations of Motion.*

We take a system (1, 2, 3) of right-handed axes of reference, see fig. 9, whose origin is O, the centre of gravity of the shell, and whose angular velocity at any instant is represented by the vector $\mathbf{\Theta}$, with components $\Theta_1$, $\Theta_2$, $\Theta_3$. The direction of the axis of the shell OA is represented by the unit vector* $\mathbf{\Lambda}$, and the direction of motion of the centre of gravity by the unit* vector $\mathbf{X}$.

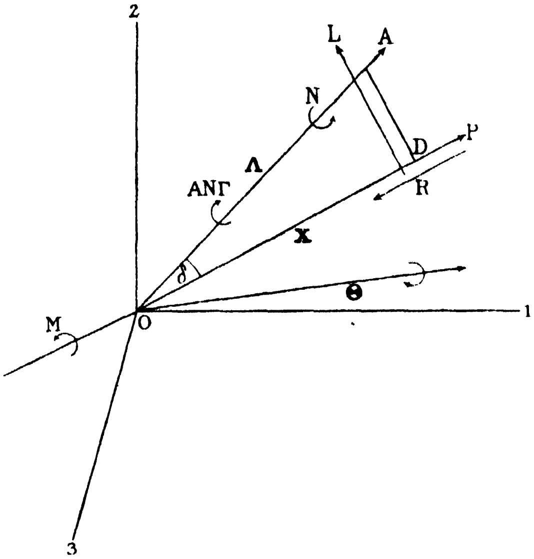


Fig. 9.

With the notation already introduced in Part I., the total angular momentum of the shell can be expressed as the sum of two vectors :—

(i.) The angular momentum about OA, $\mathrm{AN}\mathbf{\Lambda}$ ;

(ii.) The total angular momentum about a transverse axis.

If the total angular velocity about a transverse axis is $\omega$, the angular momentum is $\mathrm{B}\omega$, and is equal to the moment of momentum of a particle whose mass is B and whose distance from O is represented by the vector $\mathbf{\Lambda}$. Now the actual velocity of such a particle relative to O is $\mathbf{\Lambda}' - [\mathbf{\Lambda} . \mathbf{\Theta}]$, and therefore its moment of momentum about O is

$$\mathrm{B}\{[\mathbf{\Lambda} . \mathbf{\Lambda}'] - [\mathbf{\Lambda} . [\mathbf{\Lambda} . \mathbf{\Theta}]]\}.$$

The total angular momentum, $\mathbf{H}$, of the shell about O is therefore given by the equation

$$\mathbf{H} = \mathrm{AN}\mathbf{\Lambda} + \mathrm{B}\{[\mathbf{\Lambda} . \mathbf{\Lambda}'] - [\mathbf{\Lambda} . [\mathbf{\Lambda} . \mathbf{\Theta}]]\} ; \tag{3.101}$$

using (3.012) this becomes

$$\mathbf{H} = \mathrm{AN}\mathbf{\Lambda} + \mathrm{B}\{[\mathbf{\Lambda} . \mathbf{\Lambda}'] - (\mathbf{\Lambda} . \mathbf{\Theta})\,\mathbf{\Lambda} + \mathbf{\Theta}\}. \tag{3.102}$$

---

*I.e.* $(\mathbf{\Lambda})^2 = (\mathbf{X})^2 = 1$.

The force components that we propose to include at this stage are R, L, M, and ANΓ. To simplify the algebra we write*

$$L = \kappa m v \sin \delta, \qquad M = \mu \sin \delta.$$

The various components can then be represented by the following vectors :—

(i.) The drag R, by the vector $-\mathrm{R}\boldsymbol{\Lambda}$ ;

(ii.) The cross-wind force L, by the vector† $\kappa m v\,\{\boldsymbol{\Lambda} - \mathbf{X}\cos\delta\}$ ;

(iii.) The couple M, by the vector† $\mu\,[\mathbf{X}\,.\,\boldsymbol{\Lambda}]$ ;

(iv.) The couple ANΓ, by the vector $-\mathrm{AN}\Gamma\boldsymbol{\Lambda}$.

The complete equation for the angular motion is therefore

$$\mathbf{H}' - [\mathbf{H}\,.\,\boldsymbol{\Theta}] = \mu\,[\mathbf{X}\,.\,\boldsymbol{\Lambda}] - \mathrm{AN}\Gamma\boldsymbol{\Lambda}, \tag{3.103}$$

where $\mathbf{H}$ is given by (3.102). Taking the scalar product of both sides of (3.103) into $\boldsymbol{\Lambda}$, we obtain, with the help of (3.011)–(3.013),

$$\mathrm{N}' = -\mathrm{N}\Gamma. \tag{3.104}$$

After substituting for N′, equation (3.103), written in full, reduces to

$$\begin{aligned} \mathrm{AN}\boldsymbol{\Lambda}' + \mathrm{B}[\boldsymbol{\Lambda}\,.\,\boldsymbol{\Lambda}''] - 2\mathrm{B}(\boldsymbol{\Lambda}\,.\,\boldsymbol{\Theta})\,\boldsymbol{\Lambda}' - \mathrm{B}(\boldsymbol{\Lambda}\,.\,\boldsymbol{\Theta}')\,\boldsymbol{\Lambda} + \mathrm{B}\boldsymbol{\Theta}' \\ - \mathrm{AN}\,[\boldsymbol{\Lambda}\,.\,\boldsymbol{\Theta}] + \mathrm{B}(\boldsymbol{\Lambda}\,.\,\boldsymbol{\Theta})\,[\boldsymbol{\Lambda}\,.\,\boldsymbol{\Theta}] = \mu\,[\mathbf{X}.\boldsymbol{\Lambda}]. \end{aligned} \tag{3.105}$$

3.11. *The Equations of Motion of the Centre of Gravity.*—The velocity of the centre of gravity is represented by the vector $v\mathbf{X}$, and its acceleration is therefore represented by the vector

$$\frac{d}{dt}\{v\mathbf{X}\} - v[\mathbf{X}\,.\,\boldsymbol{\Theta}].$$

In addition to the drag and cross-wind force impressed by the air, we shall suppose that gravity is acting on the shell.

---

* The mass and velocity of the shell are $m$ and $v$ respectively. For the rest of the notation see § 1.31.

† If a perpendicular AD be drawn from A to OP, DA is parallel to the direction of the cross-wind force L, and its length is $\sin\delta$, if OA is of unity length. The vector DA is equal to the difference of the vectors OA and OD, so that it is equal to $\boldsymbol{\Lambda} - \mathbf{X}\cos\delta$. Hence $\{\boldsymbol{\Lambda} - \mathbf{X}\cos\delta\}/\sin\delta$ is the unit vector parallel to the cross-wind force. Similarly $[\mathbf{X}\,.\,\boldsymbol{\Lambda}]/\sin\delta$ is the unit vector normal to the plane AOP *i.e.*, parallel to the axis of the couple M. It is easy to verify, with the help of (3.012), that

$$\boldsymbol{\Lambda} - \mathbf{X}\cos\delta = [[\mathbf{X}\,.\,\boldsymbol{\Lambda}]\,.\,\mathbf{X}].$$

The acceleration due to gravity is represented by the vector $\mathbf{G}$, whose modulus is $g$.[*] Under these conditions the vector equation of motion of the centre of gravity is

$$(3.111)\qquad \frac{d}{dt}\{v\mathbf{X}\}-v[\mathbf{X}\,.\,\boldsymbol{\Theta}] = -\frac{\mathrm{R}}{m}\mathbf{X}+\kappa v\,\{\boldsymbol{\Lambda}-\mathbf{X}\cos\delta\}+\mathbf{G}.$$

Taking the scalar product of both sides into $\mathbf{X}$, equation (3.111) reduces to

$$(3.112)\qquad v' = -\mathrm{R}/m+(\mathbf{G}\,.\,\mathbf{X}).$$

On substituting this value of $v'$ in (3.111), and dividing by $v$, we obtain

$$(3.113)\qquad \mathbf{X}'-[\mathbf{X}\,.\,\boldsymbol{\Theta}] = \kappa\,\{\boldsymbol{\Lambda}-\mathbf{X}\cos\delta\}+\{\mathbf{G}-(\mathbf{G}\,.\,\mathbf{X})\,\mathbf{X}\}/v.$$

Equations (3.104), (3.105), (3.112), and (3.113) determine the motion completely.

### § 3.2. *Equations of Motion of Type* $\alpha$.

When a shell is initially sufficiently stable, and leaves the muzzle so that its initial disturbance is small, it will be shown[†] that the axis OA and the direction of motion OP deviate, at any time $t$, by small angles only from the direction of the tangent to the corresponding[‡] plane trajectory at the same time. This is true of the early part of all trajectories, and for the whole of a trajectory whose initial elevation is less than 45 degrees—at any rate, when the muzzle velocity is fairly large Under these circumstances we may follow the classical[§] treatment in regarding the plane trajectory as a first approximation to the actual trajectory. It is then convenient to refer the motion to axes moving with the tangent to this plane trajectory. The axis O1 is the tangent to the plane trajectory drawn in the direction of motion; axis O2 is the upward normal; and axis O3 is horizontal and to the right, as viewed from the gun. The components of $\boldsymbol{\Lambda}$ and $\mathbf{X}$ are $(l, m, n)$ and $(x, y, z)$, which are therefore the direction cosines of OA and OP respectively.

It will now be shown that it is possible to express the complete motion approximately in terms of the two complex variables, $m+in$ and $y+iz$, and the elements of the plane trajectory. We suppose that the equations of the plane trajectory have been numerically solved, so that, *e.g.*, $v_1$ and $\theta_1$, the velocity and inclination in the plane trajectory, may be regarded as tabulated functions of $t$.

---

[*] The vector $\mathbf{G}$ may, if desired, be regarded as representing any force which acts through the centre of gravity and is a function of position only.

[†] See § 4.21.

[‡] The *corresponding* plane trajectory is the trajectory which would be described by the same shell, with the same initial velocity and initial direction of motion, if its yaw remained always zero.

See, *e.g.*, CRANZ, 'Zeitschrift für Math. u. Phys.' The equations we obtain, however, appear to be new.

The components of $\boldsymbol{\Theta}$ are $(0, 0, \theta'_1)$. Using the foregoing values of the components of $\boldsymbol{\Lambda}$, $\mathbf{X}$ and $\boldsymbol{\Theta}$ in equations (3.104) and (3.105), we obtain

$$N' = -N\Gamma, \tag{3.201}$$

as before; the second and third components of (3.105) give

$$AN m' + B(nl'' - ln'') - 2Bnm'\theta'_1 - Bmn\theta''_1 + ANl\theta'_1 - Bnl\theta'^2_1 = \mu(zl - xn), \tag{3.202}$$

$$AN n' + B(lm'' - ml'') - 2Bnn'\theta'_1 - Bn^2\theta''_1 + B\theta''_1 = \mu(xm - yl). \tag{3.203}$$

To solve the equations it is necessary to neglect certain terms. A discussion of the relative magnitude of the terms neglected, in various circumstances, will be given later in § 4.3. Some of these terms are negligible in all cases, on account of the smallness of $\theta'_1$ in comparison with the angular spin of the shell. Others are only negligible so long as the yaw $\delta$ is so small that $1 - \cos\delta$ and $1 - \sin\delta/\delta$ may be neglected in comparison with unity. By such arguments it is not difficult to justify the reduction of these equations to the form

$$AN m' - Bn'' + AN\theta'_1 = \mu(z - n), \tag{3.204}$$

$$AN n' + Bm'' + B\theta''_1 = \mu(m - y). \tag{3.205}$$

For the particular case of the initial motion of the shells from the gun rifled 1 turn in 30 calibres in the present trials, the terms neglected are, in general, less than 1 per cent. of some term retained, and the coefficients of equations (3.204) and (3.205) may be regarded as affected by possible 1 per cent. errors. Even in the case of the gun rifled 1 turn in 40 calibres, where values of $\delta$ as great as 7 degrees or more are met with among the stable rounds, the employment of (3.204) and (3.205) is justifiable.

We now define new variables and constants by the equations

$$\eta + c\xi = m + in, \qquad c\xi = y + iz,$$
$$AN/B = \Omega, \qquad c = \cos\theta_1, \qquad \theta'_1 + i\theta''_1/\Omega = \Phi.$$

If we multiply (3.204) by $i$, and subtract from (3.205), we obtain

$$\frac{d^2}{dt^2}(\eta + c\xi) - i\Omega\frac{d}{dt}(\eta + c\xi) - \frac{\mu\eta}{B} = i\Omega\Phi. \tag{3.206}$$

So long as the yaw remains small, equations (3.201) and (3.206) may be taken as equivalent to (3.104) and (3.105).

3.21. *The Motion of the Centre of Gravity.*—With the present axes, the components of $\mathbf{G}$ are $(-g\sin\theta_1, -g\cos\theta_1, 0)$. Equation (3.112) becomes[†]

$$v' = -R(v, \delta)/m^* - g(x\sin\theta_1 + y\cos\theta_1). \tag{3.211}$$

---

[†] To avoid confusion the mass of the shell is temporarily denoted by $m^*$

The second and third components of (3.113) become

(3.212) $$y'+x\theta_1' = \kappa(m-y\cos\delta)-(g/v)\cos\theta_1+(yg/v)(x\sin\theta_1+y\cos\theta_1),$$

(3.213) $$z' = \kappa(n-z\cos\delta)+(zg/v)(x\sin\theta_1+y\cos\theta_1).$$

The equation of the plane trajectory corresponding to (3.211) ($\delta = y = 0$, $x = 1$) is

(3.214) $$v'_1 = -\mathrm{R}(v_1, 0)/m^*-g\sin\theta_1.$$

Therefore, if $u = v-v_1$, $u$ satisfies the equation

(3.2141) $$u' = -\{\mathrm{R}(v_1+u, \delta)-\mathrm{R}(v_1, 0)\}/m^*-g\{(x-1)\sin\theta_1+y\cos\theta_1\}.$$

In §§ 4.22, 4.31, we shall show that it is legitimate to regard the value of $u$ determined by this equation as zero. We can therefore replace $v$ by $v_1$ in (3.206), (3.212) and (3.213).

A further discussion shows that (3.212) and (3.213) can be reduced to

$$y' = \kappa(m-y)+(g/v_1)\,y\sin\theta_1,$$
$$z' = \kappa(n-z)+(g/v_1)\,z\sin\theta_1,$$

the accuracy and validity of these equations being the same as those of (3.204) and (3.205).† These equations combine to give

$$\frac{d}{dt}(c\xi) = \kappa\eta-\frac{gc\sin\theta_1}{v_1}\xi;$$

or, using the equation of the plane trajectory, $\theta'_1 = -(g/v_1)\cos\theta_1$,

(3.215) $$\xi' = \kappa\eta/c.$$

In the cases contemplated this equation is equivalent to (3.212) and (3.213). Then (3.215), (3.206) and the equations of the plane trajectory represent the required approximation to the complete equations of motion of the shell.

In order to convert (3.206) and (3.215) into linear differential equations, it is necessary to assume that $\mu$ and $\kappa$ are independent of $\delta$, and regard them as functions of $v_1$. This approximation involves errors no greater than the previous approximations. If $\Omega$ is treated as a variable, it must be determined by (3.201), $\Gamma$ being regarded as a known function of the time. All the coefficients in (3.215) and (3.206) are then known functions of the time.

### § 3.3. *Equations of Motion of Type β.*

In the neighbourhood of the vertex of a trajectory of elevation as great as 70 degrees, the yaw, as stated in § 1.34, may reach large values. In such cases, the plane trajectory can no longer be regarded as a valid first approximation, and the only possible method is to obtain equations of motion which are suitable for direct step-by-step integration. For this purpose the following set of moving axes are most suitable, as they reduce the equations of motion of the centre of gravity to its simplest form. We take the true direction of motion OP for the axis 1 and a horizontal line at right angles to OP for the axis 3. We define the position of OP by spherical polar co-ordinates $\theta$, $\psi$ with respect to axes fixed in direction at O, see fig. 10. Then $\mathbf{X}$ has components $(1, 0, 0)$, $\mathbf{\Theta}$ has components $(-\psi' \sin\theta, -\psi' \cos\theta, \theta')$, $\mathbf{G}$ has components $(-g\sin\theta, -g\cos\theta, 0)$ and $\mathbf{\Lambda}$ components $(l, m, n)$ as before.

---

† With the exception noted in § 4.22.

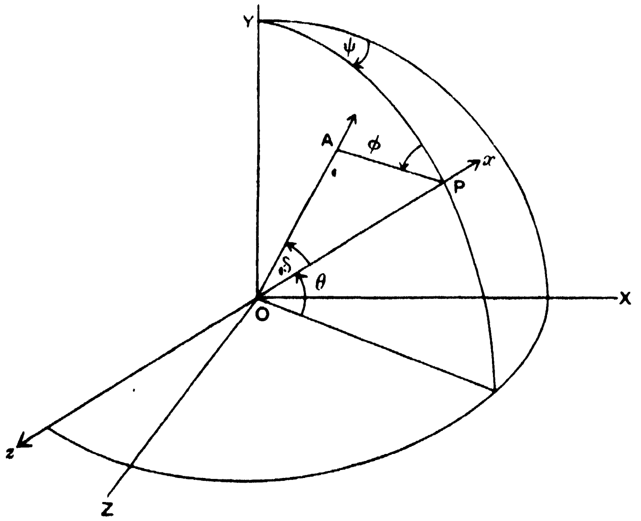


Fig. 10. OX, Y, Z are fixed axes, OY being the upwards vertical; the plane XOY contains the line of fire.

Equation (3.105), when written out in full, becomes very complicated. To simplify it, we can, under certain circumstances, neglect the angular momentum about a transverse axis compared to the angular momentum about the axis of the shell. The legitimacy of this approximation, which is equivalent to putting $B = 0$ in (3.105), is discussed in § 4.33. It should be stated that this type of approximation also is classical,* but that the equations we obtain are apparently new and of a wide range of validity.

As before, we have

$$N' = -N\Gamma. \tag{3.301}$$

The second and third components of (3.105) reduce to

$$\mathrm{AN}m' - \mathrm{AN}\left(-n\psi' \sin\theta - l\theta'\right) = -\mu n,$$
$$\mathrm{AN}n' - \mathrm{AN}\left(-l\psi' \cos\theta + m\psi' \sin\theta\right) = \mu m,$$

---

* See, *e.g.*, CHARBONNIER, 'Traité de Balistique Extérieure,' Livre V., Chap. IV.

or, writing $\omega$ for $\mu/AN$,

(3.302) $$m' = -n\,(\omega+\psi'\sin\theta)-l\theta',$$

(3.303) $$n' = m\,(\omega+\psi'\sin\theta)-l\psi'\cos\theta.$$

The corresponding equations of motion of the centre of gravity are

(3.304) $$v' = -R\,(v,\,\delta)/m^{*}-g\sin\theta,$$

(3.305) $$\theta' = \kappa m-(g/v)\cos\theta,$$

(3.306) $$\psi'\cos\theta = \kappa n.$$

The six equations (3.301) to (3.306) can be solved by a step-by-step process, if R, $\kappa$, $\mu$ and $\Gamma$ are numerically known functions of $v$ and $\delta$. They are valid without restriction as to the size of $\delta$, and have proved of value for the discussion of trajectories at very high elevations. They are, however, necessarily invalid when any question of stability is under discussion.

## § 3.4. *Equations of Motion of Type $\gamma$.*

For the purpose of discussing the initial motion of a shell which is unstable or just stable, equations of types $\alpha$ and $\beta$ are invalid, and it is necessary to make use of equations corresponding to the equations of energy and angular momentum for a top. The equations we shall thus obtain are of far less general applicability than types $\alpha$ and $\beta$.

With this object we take the scalar product of both sides of equation (3.105) into the vector $[\mathbf{\Lambda}\,.\,\mathbf{\Lambda}']+\mathbf{\Theta}$, and obtain, after reduction,

(3.401) $$\tfrac{1}{2}B\frac{d}{dt}\left\{(\mathbf{\Lambda}')^2+2\,(\mathbf{\Theta}\,.\,[\mathbf{\Lambda}\,.\,\mathbf{\Lambda}'])+(\mathbf{\Theta})^2-(\mathbf{\Lambda}\,.\,\mathbf{\Theta})^2\right\} = -\mu\,(\mathbf{X}\,.\,\{\mathbf{\Lambda}'-[\mathbf{\Lambda}\,.\,\mathbf{\Theta}]\}).$$

Using the axes described in the last section, we note that, over a limited range at the beginning of a trajectory, the first two components of $\mathbf{\Theta}$ are numerically very small compared to the third, $\theta'$. We shall find that the effect of $\theta'$ itself is negligible in the cases we consider. We shall therefore neglect the other components of $\mathbf{\Theta}$ at once. Taking $\delta$ and $\phi$ as spherical polar co-ordinates of the axis OA referred to the moving axes, so that

$$l = \cos\delta,\ m = \sin\delta\cos\phi,\ n = \sin\delta\sin\phi,\dagger$$

---

† The angle $\phi$ is not exactly the angle measured by the jump cards, but the difference is negligible. The angle $\delta$ is exactly the measured angle of yaw.

we find that (3.401) reduces to

$$(3.402)\quad \tfrac{1}{2}\mathrm{B}\frac{d}{dt}\left\{\delta'^2+\phi'^2\sin^2\delta+2\theta'(\delta'\cos\phi-\phi'\sin\delta\cos\delta\sin\phi)+\theta'^2(1-\sin^2\delta\sin^2\phi)\right\}$$
$$= -\mu\left\{\frac{d\cos\delta}{dt}-\theta'\sin\delta\cos\phi\right\}.$$

This is the equation corresponding to the equation of energy. The first component of (3.105) corresponds to the equation of angular momentum for a top, and reduces in the same way to

$$(3.403)\quad \frac{d}{dt}\{\mathrm{AN}\cos\delta+\mathrm{B}\phi'\sin^2\delta\}+2\mathrm{B}\theta'\delta'\sin^2\delta\sin\phi-\mathrm{B}\theta''\sin\delta\cos\delta\sin\phi$$
$$-\mathrm{AN}\theta'\sin\delta\cos\phi+\mathrm{B}\theta'^2\sin^2\delta\cos\phi\sin\phi = 0.$$

Equation (3.201) remains unaltered. Over the range of the jump card experiments a mean value of $\delta'/\theta'$ is 50. We shall therefore regard it as legitimate for our present purposes to neglect all terms containing $\theta'$. On integrating the resulting equations we obtain

$$(3.404)\qquad \tfrac{1}{2}\mathrm{B}(\delta'^2+\phi'^2\sin^2\delta)+\int_0^\delta \mu d\cos\delta = \tfrac{1}{2}\mathrm{BE},$$

$$(3.405)\qquad \mathrm{AN}\cos\delta+\mathrm{B}\phi'\sin^2\delta = \mathrm{BF},$$

where E and F are constants of integration. In (3.405) it is assumed that N is constant. If $\mu$ is constant these equations are of the same form as those of the motion of a top. In the more important applications to the jump card trial which we shall make of (3.404) and (3.405), $\mu$ will be treated as a variable function of $\delta$, and also of $v$.

### § 3.5. *The Additional Force Components* H, J *and* K.

It is now necessary to consider the effect of the additional force components, mentioned in §§ 1.1, 1.12, and denoted by H, J and K. These have so far been omitted from the general equations for the sake of simplicity. The couples H and J will affect the angular motion of the axis, and the force K will affect the motion of the centre of gravity. For algebraic convenience we define new variables $h$, $\gamma$, $\lambda$, by the equations

$$\mathrm{H} = h\mathrm{B}w,\qquad \mathrm{J} = \mathrm{AN}\gamma\sin\delta,\qquad \mathrm{K} = m\mathrm{N}v\lambda\sin\delta,$$

where $w$ is the total angular velocity of the axis of the shell. The force components may then be represented, in the notation of § 3.1, by the following vectors :—

H by the vector $-h\mathrm{B}\{[\mathbf{\Lambda}\,.\,\mathbf{\Lambda}']-(\mathbf{\Lambda}\,.\,\mathbf{\Theta})\,\mathbf{\Lambda}+\mathbf{\Theta}\}$ ;

J by the vector $\mathrm{AN}\gamma\,(\mathbf{\Lambda}\cos\delta-\mathbf{X})$ ;

K by the vector $m\mathrm{N}v\lambda\,[\mathbf{\Lambda}\,.\,\mathbf{X}]$.

To include the effect of these components we add to the right-hand side of (3.105)

$$-h\mathrm{B}\{[\mathbf{\Lambda}\,.\,\mathbf{\Lambda}']-(\mathbf{\Lambda}\,.\,\mathbf{\Theta})\,\mathbf{\Lambda}+\mathbf{\Theta}\}+\mathrm{AN}\gamma\,(\mathbf{\Lambda}\cos\delta-\mathbf{X}),$$

and to the right right-hand side of (3.113)

$$\mathrm{N}\lambda\,[\mathbf{\Lambda}\,.\,\mathbf{X}].$$

Equations (3.104) and (3.112) (type $\alpha$) are unaltered. As a result, the following additions must be made to the right-hand side of succeeding equations :—

To (3.202), $+h\mathrm{B}n'+\mathrm{AN}\gamma\,(m-y)$.
To (3.203), $-h\mathrm{B}m'+\mathrm{AN}\gamma\,(n-z)$.
To (3.212), $-\mathrm{N}\lambda\,(z-n)$.
To (3.213), $-\mathrm{N}\lambda\,(m-y)$.

As the total effect of the extra components $h$, $\gamma$ and $\lambda$ is certainly small in any practical case, we have neglected all terms other than those of the lowest order in $\delta$. Equations (3.206) (3.215), when modified by the inclusion of these extra terms, become

$$\frac{d^2}{dt^2}(\eta+c\xi)-(i\Omega-h)\frac{d}{dt}(\eta+c\xi)-\left(\frac{\mu}{\mathrm{B}}-i\Omega\gamma\right)\eta = i\Omega\Phi\,; \tag{3.501}$$

$$\xi' = (\kappa-i\mathrm{N}\lambda)\,\eta/c. \tag{3.502}$$

3.51. *The Additional Terms in Equations of Types $\beta$ and $\gamma$.*—The additional terms in the equations of type $\beta$ can be written down in a similar manner. The following additions must be made to the right-hand sides of the equations :—

To (3.302), $+\gamma m\cos\delta-h\,(nl'-ln')/\Omega$.
To (3.303), $+\gamma n\cos\delta-h\,(lm'-ml')/\Omega$.
To (3.305), $+\mathrm{N}\lambda n$.
To (3.306), $-\mathrm{N}\lambda m$.

The terms in $h$ are negligible, as they are $\mathrm{O}\,(h\delta'/\Omega^2)$ compared with the principal terms $-n\omega-l\theta'$, so long as $\omega/\Omega$ is not very small. The principal application of these equations is to the motion of a shell near the vertex of a trajectory at an elevation of 70 degrees, where the velocity becomes small while the spin probably remains large. Under these circumstances the terms $\gamma$ and $\lambda$ arising from the spin rise in importance relatively to the terms $\omega$ and $\kappa$ representing the ordinary force components. The inclusion of the extra terms $\gamma$ and $\lambda$ in these equations is at present of no practical importance, as we have no definite information as to their value.

The corresponding terms could be added to equations of type $\gamma$ by the same methods, but the results are of no importance as it is impossible at present to solve these equations unless these terms are neglected.

### § 3.6. *The Approximate Solution of Equations of Type $\alpha$.*

3.61. *The Nature of the Solution Required.*—The system of equations (3.501), (3.502) are linear differential equations with respect to the time of the second order, the coefficients being regarded as known functions of the time $t$. Since these functions are in practice empirical and by no means simple, an exact solution is impossible. To simplify the discussion we write

$$s = A^2N^2/4B\mu, \qquad \mu/B = \Omega^2/4s,$$

so that equations (3.501), (3.502) become

$$(3.611) \qquad \frac{d^2}{dt^2}(\eta+c\xi)-(i\Omega-h)\frac{d}{dt}(\eta+c\xi)-\left(\frac{\Omega^2}{4s}-i\Omega\gamma\right)\eta = i\Omega\Phi\,;$$

$$(3.612) \qquad \xi'-(\kappa-iN\lambda)\,\eta/c = 0.$$

If the terms in $\xi$, $h$, $\gamma$ be omitted from (3.611), and $s$, N and $\Omega$ are assumed constant, the equation reduces to that for the small oscillations of a top in the neighbourhood of the vertical.

The coefficient $s$ is the stability factor as defined in § 1.31. In order to be able to apply the approximations on which (3.611) and (3.612) are based, we shall find that it is necessary to assume that the shell is more than just stable, *e.g.*, $s>1{\cdot}1$.

We proceed to develop an approximate solution of the equations on the assumption that $\Omega$ is large. If we ignore the dimensions of the various terms, and take the unit of time as 1 second, then $\Omega$ is in practice greater than 100 (radians per second), all other terms being of the order unity. This is really equivalent to assuming that all the ratios $\kappa/\Omega$, $h/\Omega$, ..., which are of no dimensions, are small. It will be found necessary to assume further that all derivatives with respect to the time are of order unity in units of 1 second, *e.g.*, that $\kappa'$, $\kappa''$, $s'$, $\Omega'$..., are of order unity. These conditions are satisfied in practice. As a result, we can say that $\kappa/\Omega$, $s'/\Omega$..., are small quantities of the first order, and $\kappa'/\Omega^2$, $s''/\Omega^2$, $\theta''_1/\Omega^2$..., are small quantities of the second order. For simplicity, we shall throughout ignore dimensions, and denote such terms of the first order by $O(1/\Omega)$, and terms of the second order by $O(1/\Omega^2)$.*
The arithmetical values of the various terms are investigated in detail in § 4.3 below.

The above facts indicate the lines on which an approximate solution is to be sought—*we require the asymptotic expansion of the solution (or its leading terms) for large values of the parameter $\Omega$.* Methods of obtaining such expansions have been investigated in general terms by **Horn** and **Schlesinger.**[*] A method, which is slightly different algebraically, is more convenient here; the asymptotic properties of our solutions, however, may be regarded as established by the researches of these authors.

---

* In practice the spin N, and therefore $\Omega$, decreases slightly along the trajectory, but the diminution is not sufficient to affect the assumption that $\Omega$ is large.

The equations (3.611) and (3.612) are a pair of linear differential equations with respect to the time for the two dependent variables $\eta$ and $\zeta$, (3.611) being of the second order and (3.612) of the first. There must, therefore, be three independent solutions.

It is convenient to eliminate $\zeta'$ and $\zeta''$ from (3.611) by the use of (3.612), the result being

$$(3.613) \qquad \eta''-(i\Omega-h-\kappa_1)\,\eta'-\{\Omega^2/4s+i\Omega\,(\kappa_1-\gamma)-h\kappa_1-\kappa'_1-\kappa_1c'/c\}\,\eta \\ -\{i\Omega c'-hc'-c''\}\,\zeta=i\Omega\Phi,$$

where $\kappa_1$ is written for $(\kappa-i\mathrm{N}\lambda)$. It is believed that $\mathrm{N}\lambda$ is small compared with $\kappa$, so that for simplicity the term $\mathrm{N}\lambda$ will usually be omitted in subsequent work. The term $\gamma$ will however be retained.

3.62. *The Complementary Function.*—A first approximation to the three independent complementary functions is obtained, following Horn and Schlesinger, by making the substitution,

$$\eta=e^{i\Omega x}\bar{\eta}, \qquad \zeta=e^{i\Omega x}\bar{\zeta},$$

and treating $\bar{\eta}$ and $\bar{\zeta}$ as constants in determining $\eta'$ and $\zeta'$. We also neglect all but the highest order terms in $\Omega$ in each equation. The equations then reduce to

$$(3.621) \qquad (-\Omega^2x'^2+\Omega^2x'-\Omega^2/4s)\,\bar{\eta}-i\Omega c'\bar{\zeta}=0\,;$$

$$(3.622) \qquad -\kappa\bar{\eta}/c+i\Omega x'\bar{\zeta}=0.$$

On eliminating $\bar{\eta}$ and $\bar{\zeta}$, and retaining only the terms of highest order in $\Omega$, these reduce to

$$x'\,(x'^2-x'+1/4s)=0,$$

a cubic equation for $x'$ whose three roots correspond to the three independent solutions required. The roots are $\frac{1}{2}\pm\frac{1}{2}(1-1/s)^{\frac{1}{2}}$, 0, or writing, for shortness, $\sigma = (1-1/s)^{\frac{1}{2}}$, the three values of $x$ are

---

[*] J. Horn, 'Mathematische Annalen,' vol. 52, p. 271 and p. 340. L. Schlesinger, *ibid.*, vol. 63, p. 277; 'Comp. Rend.,' vol. 142, p. 1031. The investigations of the complementary function given by these writers are fairly complete, the asymptotic nature of the expansions being established. The latter writer considers a system of $n$ linear differential equations. A similar treatment of the complementary function and the particular integral of a special equation is suggested (without proof) by M. de Sparre 'Atti (Rendiconti) della R. Acc. dei Lincei,' 1898, Ser. V., vol. 72, p. 111; this writer was obviously led to the solution he gives by his researches on the motion of spinning projectiles.

[*Note added July* 30, 1920. See also G. D. Birkhoff, 'Trans. Amer. Math. Soc', vol. 9, p. 219.]

$$x_1 = \tfrac{1}{2}\int_0^t (1+\sigma)\,dt: \qquad x_2 = \tfrac{1}{2}\int_0^t (1-\sigma)\,dt; \qquad x_3 = 0.$$

It appears that the first two solutions correspond to the complementary function of equation (3.613) with the term in $\zeta$ neglected, so that $\eta$ is large compared with $\zeta$. If $s < 1$, $\sigma$ is imaginary, the motion is unstable and the solution fails. In the third solution $\zeta$ is large compared with $\eta$, and a first approximation to it gives a constant value to $\zeta$, obtained by neglecting the term in $\eta$ in equation (3.612). It is convenient to obtain the first two solutions independently by a special method.

We first omit the term in $\zeta$ in equation (3.613); it is not required till the second approximation. Write the equation, for simplicity, in the form

(3.623)
$$\eta''-i\mathrm{A}\eta'-\mathrm{B}\eta = 0,$$

where

$$\mathrm{A} = \Omega+ih+i\kappa,$$
$$\mathrm{B} = \Omega^2/4s+i\Omega(\kappa-\gamma)-h\kappa-\kappa'-\kappa c'/c.$$

Remove the second term by substituting

$$\eta = y\exp\left\{\tfrac{1}{2}i\int_0^t \mathrm{A}\,dt\right\},$$

giving

(3.6231)
$$y''+\mathrm{M}y = 0,$$

where

$$\begin{aligned}\mathrm{M} &= \tfrac{1}{4}\mathrm{A}^2-\mathrm{B}+\tfrac{1}{2}i\mathrm{A}' \\ &= \tfrac{1}{4}\Omega^2\sigma^2\left\{1+\frac{2i}{\Omega\sigma^2}(h-\kappa+2\gamma+\mathrm{N}'/\mathrm{N})+\mathrm{O}\left(\frac{1}{\Omega^2}\right)\right\}.\end{aligned}$$

Substitute $y = \mathrm{R}e^{\pm i\mathrm{P}}$, so that (3.6231) becomes

(3.6232)
$$\mathrm{R}''\pm(2i\mathrm{P}'\mathrm{R}'+i\mathrm{P}''\mathrm{R})-\mathrm{P}'^2\mathrm{R}+\mathrm{MR} = 0.$$

We may make P and R satisfy any single relation we choose, *e.g.*,

$$2\mathrm{P}'\mathrm{R}'+\mathrm{P}''\mathrm{R} = 0,$$

giving $\mathrm{P}' = 1/\mathrm{R}^2$,* so that (3.6232) becomes

(3.6233)
$$\mathrm{R}''-1/\mathrm{R}^3+\mathrm{MR} = 0.$$

---

* More generally $\mathrm{P}' = a/\mathrm{R}^2$, where $a$ is a constant, but the value of this constant is immaterial, as it disappears in the result.

This equation may be solved asymptotically, by successive approximation, by writing*

$$R = R_0(1+R_1+R_2\ldots),$$

where

$$R_1 = O(1/M) = O(1/\Omega^2), \qquad R_2 = O(1/\Omega^4)\ldots.$$

We obtain, in succession, the approximations

$$R_0 = M^{-\frac{1}{4}},$$

$$R_1 = -\tfrac{1}{4}M^{-\frac{3}{4}}\frac{d^2}{dt^2}(M^{-\frac{1}{4}}),$$

verifying the relation $R_1 = O(1/\Omega^2)$. The order of magnitude of $R_1$ in practice will be discussed in § 4.32, where it will be shown to be negligible.† We therefore take as our two standard solutions

$$y_1 = M^{-\frac{1}{4}}\exp\left\{i\int_0^t M^{\frac{1}{2}}\,dt\right\},$$

$$y_2 = M^{-\frac{1}{4}}\exp\left\{-i\int_0^t M^{\frac{1}{2}}\,dt\right\},$$

giving, for the complementary function of (3.623),

$$\eta = (\Omega\sigma)^{-\frac{1}{2}}\{1+O(1/\Omega)\}\{K_1e^{iP_1}+K_2e^{iP_2}\}, \tag{3.6234}$$

where $K_1$, $K_2$ are arbitrary constants, and $P_1$, $P_2$ are given by

$$P_1,\ P_2 = \tfrac{1}{2}\int_0^t[\Omega(1\pm\sigma)+i\{h+\kappa\pm(h-\kappa+2\gamma+N'/N)/\sigma\}]\,dt.$$

This is the form of solution which is used in analysing the jump card experiments, and contains all the terms that can be required in practice.

It is now necessary to examine the effect of the term in $\zeta$ in (3.613), which has so far been omitted. The value of $\zeta'$, obtained from (3.612), corresponding to the first solution for $\eta$, is

$$\zeta' = (\kappa/c)(\Omega\sigma)^{-\frac{1}{2}}e^{iP_1},$$

so that, on integrating by parts to obtain the leading terms,

$$\zeta_1 = \frac{2\kappa\eta_1}{ic\Omega(1+\sigma)}+O\left(\frac{1}{\Omega^2}\right).$$

---

* At this point the advantage of our *ad hoc* method over more general methods is apparent, as we obtain in one step a solution with an error $O(1/\Omega^2)$, whereas the general method requires two steps.

† We assume that the numerical value of $R_1$, the next term in the expansion, is a measure of the error in the solution caused by omitting all terms after the first. The expansions for $y_1$ and $y_2$ are known to be asymptotic for large values of $\Omega$, so that the error will be some finite multiple of $R_1$, but the size of the numerical factor is unknown.

Similarly we have

$$\zeta_2 = \frac{2\kappa\eta_2}{ic\Omega(1-\sigma)} + O\left(\frac{1}{\Omega^2}\right),$$

verifying that $\zeta$ is small compared to $\eta$. The contribution of $\zeta$ to equation (3.613) is thus

$$(3.6235) \qquad -\frac{2\kappa c'\eta}{c(1\pm\sigma)} + O\left(\frac{1}{\Omega}\right),$$

which is equivalent to an addition to the coefficient of $\eta$ of terms which are $O(1/\Omega^2)$ compared to the principal term. The solution can be repeated with these terms included, but is unaffected to the order to which we are working. We shall take our first two standard solutions in the form

$$(3.624) \qquad \eta_1 = \left(\frac{\Omega_0\sigma_0}{\Omega\sigma}\right)^{\frac{1}{2}} e^{iP_1}, \qquad \zeta_1 = \frac{2\kappa\eta_1}{ic\Omega(1+\sigma)},$$

$$(3.625) \qquad \eta_2 = \left(\frac{\Omega_0\sigma_0}{\Omega\sigma}\right)^{\frac{1}{2}} e^{iP_2}, \qquad \zeta_2 = \frac{2\kappa\eta_2}{ic\Omega(1-\sigma)}.$$

The differential coefficients of the solutions may be obtained by differentiation of these equations.

For the third solution we have shown that the exponential index is zero to the first order, and that a first approximation is given by

$$\eta_3 = \zeta'_3 = 0, \qquad \zeta_3 = 1.$$

The expansions take a somewhat different form, like those for the particular integral, and we write

$$(3.6261) \qquad \eta_3 = \eta^{(0)} + \eta^{(1)}/i\Omega + \eta^{(2)}/(i\Omega)^2 \ldots,$$

$$(3.6262) \qquad \zeta_3 = \zeta^{(0)} + \zeta^{(1)}/i\Omega + \zeta^{(2)}/(i\Omega)^2 \ldots.$$

Substituting in equations (3.612) and (3.613), we obtain

$$\eta^{(1)} = 4sc' = -4s\,\theta'_1 \sin\theta_1,$$

$$\zeta^{(1)} = 4\int_0^t \kappa sc'dt/c.$$

The significance of this solution will be considered after the particular integral has been discussed.

Our standard third solution is then

$$(3.627) \qquad \eta_3 = \frac{4sc'}{i\Omega}, \qquad \zeta_3 = 1 + \frac{4}{i\Omega}\int_0^t \frac{\kappa sc'}{c}\,dt.$$

3.63. *The Particular Integral.*—The question of the particular integral is not treated generally by HORN and SCHLESINGER. The former considers shortly a very particular case.* Their methods can, however, be extended to obtain the results we require.

We assume an expansion for the particular integral $\bar{\eta}$, $\bar{\zeta}$ of the form of (3.6261), (3.6262). This integral can be specified in such a way that initially $\bar{\zeta} = 0$, *i.e.*, $\zeta^{(0)} = \zeta^{(1)} = \ldots = 0$, and $\eta^{(0)} = 0$. It will then be found to be unique.

On substituting in equations (3.612) and (3.613), with the right-hand side retained, and equating powers of $i\Omega$, we find first that $\zeta^{(0)} = \eta^{(0)} = 0$ for all time, and then

$$(3.631) \qquad \eta^{(1)} = 4s\Phi, \qquad \zeta^{(1)} = \int_0^t 4s\kappa\Phi dt/c.$$

The first two terms in the expansions of $\bar{\eta}$ and $\bar{\zeta}$ take the forms

$$(3.632) \qquad \bar{\eta} = \frac{4s\Phi}{i\Omega} + \frac{4s}{(i\Omega)^2}\left\{\frac{d}{dt}(4s\Phi) + (\kappa-\gamma)\,4s\Phi + c'\int_0^t 4s\kappa\Phi dt/c\right\};$$

$$(3.633) \qquad \bar{\zeta} = \int_0^t \kappa\bar{\eta} dt/c.$$

Equations (3.632) and (3.633) will be taken as the standard particular integral. Since, moreover, they contain no periodic terms, and the initial value of $\bar{\zeta}$ is zero and those of $\bar{\eta}$ and $\bar{\eta}'$ very small in practice, it is convenient to take this solution as the standard solution of the equations of motion in cases where the initial values of $\eta$ and $\eta'$ are not exactly known—*e.g.*, in calculating the drift.

The expansions for $\bar{\eta}$ and $\bar{\zeta}$, of which the first two terms are given above, can be shown to be asymptotic, but we cannot take up this question here. The numerical accuracy of (3.632) and (3.633) will be considered in § 4.33.

3.64. The general solution of (3.612) and (3.613) may be put in the form

$$(3.641) \qquad \eta = K_1\eta_1 + K_2\eta_2 + K_3\eta_3 + \bar{\eta},$$

$$(3.642) \qquad \zeta = K_1\zeta_1 + K_2\zeta_2 + K_3\zeta_3 + \bar{\zeta},$$

where $K_1$, $K_2$, $K_3$ are arbitrary complex constants and $\eta_1, \ldots, \zeta_1, \ldots$, have the values determined in the last section.

The particular integral $\bar{\eta}$, $\bar{\zeta}$ represents the motion in an actual trajectory in which $\zeta$ is initially zero, and $\eta$ and $\eta'$ start with what may be called their equilibrium values, which are numerically very small. The solution $(K_3\eta_3+\bar{\eta})$, $(K_3\zeta_3+\bar{\zeta})$ represents the motion, of the same type, in a trajectory whose initial tangent makes an angle (determined by $K_3$) with the initial tangent of the chosen plane trajectory. This can be seen from the following considerations. The motion in a slightly different *plane trajectory* would be obtained by omitting all terms in $\eta$ from the equations of motion of the centre of gravity, and ignoring the equations of angular motion. Equation (3.612) then reduces to $\zeta' = 0$; this represents a trajectory which only differs from a varied *plane trajectory* on account of terms omitted in § 3.21, whose retention renders the equation non-linear. The value of $\eta_3$ in (3.627) gives the alteration, through the change in direction of projection, of the first term in $\bar{\eta}$.

---

* *Loc. cit.*, p. 340. We hope to publish these extensions in another place.

3.65. In the usual practical case, the initial conditions take the form

$$\zeta_0 = 0, \qquad \eta_0 = a, \qquad \eta'_0 = b\Omega,$$

where $a$ and $b$ are arbitrary complex constants. It is desirable in such a case to know the degree of importance of the three standard solutions.

The initial values of the standard solutions (retaining the highest order terms only) are as follows:—

$$\begin{array}{lll} \eta_1 = 1, & \zeta_1 = O(1/\Omega), & \eta'_1 = \tfrac{1}{2}i\Omega(1+\sigma), \\ \eta_2 = 1, & \zeta_2 = O(1/\Omega), & \eta'_2 = \tfrac{1}{2}i\Omega(1-\sigma), \\ \eta_3 = O(1/\Omega), & \zeta_3 = 1, & \eta'_3 = O(1/\Omega), \\ \bar{\eta} = O(1/\Omega), & \bar{\zeta} = 0, & \bar{\eta}' = O(1/\Omega). \end{array}$$

The constants $K_1$, $K_2$, $K_3$ are determined by the equations

$$\begin{aligned} K_1\eta_1 + K_2\eta_2 + K_3\eta_3 + \bar{\eta} &= a, \\ K_1\eta_1' + K_2\eta_2' + K_3\eta_3' + \bar{\eta}' &= b\Omega, \\ K_1\zeta_1 + K_2\zeta_2 + K_3\zeta_3 + \bar{\zeta} &= 0. \end{aligned} \tag{3.6501}$$

Retaining only the highest order terms these reduce to

$$K_1 + K_2 = a, \tag{3.651}$$

$$\tfrac{1}{2}i(1+\sigma)K_1 + \tfrac{1}{2}i(1-\sigma)\kappa_2 = b, \tag{3.652}$$

$$K_3 + O(1/\Omega) = 0. \tag{3.653}$$

It follows at once that $K_3\eta_3$ is completely negligible compared to $K_1\eta_1$ and $K_2\eta_2$, and that in investigating $\eta$ we may ignore the third solution (and the particular integral) altogether. On the other hand, the contributions of all the solutions to $\zeta$ are of the same order of importance. We shall therefore take as the solution satisfying the most general initial conditions—

$$\eta = K_1\eta_1 + K_2\eta_2, \tag{3.654}$$

where $K_1$ and $K_2$ are determined by (3.651) and (3.652), and

(3.655) $$\zeta = K_1\zeta_1 + K_2\zeta_2 + K_3\zeta_3 + \int_0^t \kappa\bar{\eta}\,dt/c,$$

where by (3.6501),

(3.656) $$K_3 = -K_1(\zeta_1)_0 - K_2(\zeta_2)_0.$$

## § 3.7. *The Solution of Equations of Type $\gamma$.*

The equations of type $\beta$ are only soluble numerically by step-by-step integration, and will not be considered here, but the equations of type $\gamma$ (§ 3.4) reduce, when $\mu$ is constant and damping effects are neglected, to the equations of a spinning top, and it is convenient to summarize here their solution, in terms of elliptic functions, in the form which is most suitable for our purposes. We shall only consider the initial conditions $\delta = 0$, $\delta' = b\Omega$; this is the rosette form of motion (§ 1.3) and is usually a good approximation to the true motion in its earliest stage. In this case we obtain from (3.404) and (3.405)

(3.701) $$\phi' = \Omega/(1 + \cos\delta),$$

(3.702) $$\delta'^2 \sin^2\delta - \Omega^2 b^2 \sin^2\delta + \Omega^2(1 - \cos\delta)^2 - (\Omega^2/2s)(1 - \cos\delta)\sin^2\delta = 0.$$

If we take $\Omega t$ as independent variable, the motion depends only on two parameters, $b$ and $s$. The solution of (3.702) is given by

(3.703) $$\sin\tfrac{1}{2}\delta = \sin\tfrac{1}{2}\alpha\ \mathrm{cn}\,(K - \lambda\Omega t,\ k),$$

where $\alpha$, $\lambda$, and $k$ are given by the formulæ

(3.7041) $$\sqrt{s} = \cos\tfrac{1}{2}\alpha \cosh\tfrac{1}{2}c,$$

(3.7042) $$b = \tan\tfrac{1}{2}\alpha \tanh\tfrac{1}{2}c,$$

(3.7043) $$\tan\epsilon = \sin\tfrac{1}{2}\alpha/\sinh\tfrac{1}{2}c, \qquad (k = \sin\epsilon),$$

(3.7044) $$\lambda = (\sin\tfrac{1}{2}\alpha)/2k\sqrt{s},$$

and K is the complete elliptic integral of the first kind to modulus $k$. Thus the yaw oscillates between the values 0 and $\alpha$, and the value of the period T—the interval between successive zeros—is given by

(3.705) $$\Omega T = 2K/\lambda.$$

The curve of yaw, $\delta$, plotted against $\Omega t$ is initially concave (convex) upwards, when $s < 1$ ($> 1$). This corresponds to the case of instability (stability) for small oscillations.

In the practical analysis of the results of rounds whose stability factor is less than or near 1, it is convenient to use graphical methods. If the observed yaw is plotted against $\Omega t$, it is easy to read off the observed values of $\alpha$, the maximum yaw, and $\Omega T$, the period. A chart was therefore constructed with suitable families of curves, according to (3.7041)–(3.7044), from which, when $\Omega T$ and $\alpha$ are known, $s$, $b$, $c$, and $k$ can be read off directly.

## PART IV.—ANALYSIS OF THE EXPERIMENTAL RESULTS.

### § 4.0. *Equations of Motion in Polar Co-ordinates.*

The theoretical results of Part III. will now be applied to the analysis of the observations described in Part II., which consist of determinations of yaw $\delta$ and orientation of yaw $\phi$, for a shell fired horizontally over a range of about 600 feet. When the stability factor is greater than about 1·1, the maximum yaw for the corresponding round never exceeds 7 degrees, and it is then possible to make use of the complementary function solution of equations of type $\alpha$ as given in § 3.6. These rounds give more valuable information than those which are less stable.

We treat certain of the force coefficients as constants over the range of the experiments, and verify that the results of the theory agree with experiment when certain values are given to the force coefficients. In particular the spin is treated as constant. The way in which the coefficients vary with the velocity is determined mainly by firing shells with various muzzle velocities. The final results have been already described in § 1.2 above.

The experiments determine the values, at definite time intervals along the range (§ 2.0), of the angle of yaw $\delta$ and the angle $\phi$ turned through by the line in which the plane of yaw meets the cards. The measured value of $\phi$ is zero, when this line is vertical and increases from 0 to $2\pi$ radians in the direction in which the shell is spinning. It is, of course, ambiguous by an integral multiple of $2\pi$. Except where specially stated the yaw $\delta$ is assumed to be an essentially positive quantity. When OA passes through the position OP, the yaw vanishes; the value of $\phi$ will change discontinuously by an amount $\pm\pi$, and $d\delta/dt$ will change its sign discontinuously.

It is convenient in Part IV. to express the solution of the equations of motion of type $\alpha$ in terms of the co-ordinates $\delta$ and $\phi$. The exact relations between the measured $\delta$ and $\phi$ and the direction cosines $(l, m, n)$ and $(x, y, z)$ of § 3.2 are

$$\cos\delta = lx+my+nz,$$

$$\tan\phi = \frac{(nx-lz)\cos\theta_1-(ny-mz)\sin\theta_1}{mx-ly},$$

where $\theta_1$ is the inclination to the horizontal of the tangent to the plane trajectory. Since $\theta_1 < 1\frac{1}{2}$ degrees, we may replace the latter by

$$\tan\phi = (nx-lz)/(mx-ly).$$

Since $\eta$ is defined by the equation

$$\eta = (m-y)+i(n-z),$$

we obtain, when $\delta$ is sufficiently small,

$$\eta = \sin\delta e^{i\phi};$$

this expression neglects terms of the second order compared to those retained. It is an adequate approximation provided $\delta < 7$ degrees.

The general solution for the equations of type $\alpha$, given in § 3.65, equations (3.654) and (3.6234), is*

$$(4.01) \qquad \eta = (\sigma_0/\sigma)^{\frac{1}{2}}(\mathrm{K}_1 e^{i\mathrm{P}_1}+\mathrm{K}_2 e^{i\mathrm{P}_2}),\dagger$$

if we ignore, as we may, the particular integral and the third solution. We shall write

$$\mathrm{P}_1 = p_1+iq_1+p_2+iq_2,$$
$$\mathrm{P}_2 = p_1+iq_1-(p_2+iq_2).$$

Then

$$(4.011) \qquad p_1 = \tfrac{1}{2}\int_0^t \Omega\,dt = \tfrac{1}{2}\Omega t, \qquad p_2 = \tfrac{1}{2}\int_0^t \Omega\sigma\,dt,$$

$$(4.012) \qquad q_1 = \tfrac{1}{2}\int_0^t (h+\kappa)\,dt, \qquad q_2 = \tfrac{1}{2}\int_0^t (h-\kappa+2\gamma)\,dt/\sigma,$$

and $\sigma^2 = 1-1/s$. We observe that $p_1$, $p_2$, $q_1$, $q_2$ are all nearly proportional to the time $t$.

The general solution (equation (4.01)) contains two complex arbitrary constants or four real constants. By a suitable choice of origin for $t$ and $\phi$ these may be reduced to two. If the time $t = 0$ corresponds to a minimum of $\delta$ and the value $\phi = 0$, equation (4.01) may be written

$$(4.02) \qquad \eta = \mathrm{J}\,(\sigma_0/\sigma)^{\frac{1}{2}} e^{ip_1-q_1}\{\cos p_2 \sinh(j-q_2)+i\sin p_2 \cosh(j-q_2)\},$$

---

* Treating N and $\Omega$ as constant, *i.e.*, neglecting the spin reducing couple $\Gamma$.

† Equation (4.01) reduces approximately to the form $\eta = \mathrm{K}t$, when $s = 1$, and to the form $\eta = (\sigma_0/\sigma)^{\frac{1}{2}}\{\mathrm{K}_1 e^{\phi_1}+\mathrm{K}_2 e^{\phi_2}\}$, when $s < 1$, and the shell unstable, the principal parts of $\phi_1$, $\phi_2$ being real and positive. The solution then fails completely as an approximation to the actual motion except over a small part of the first period. As $s$ approaches the value unity from above, the errors from this cause will begin to increase, but the magnitude of these errors can only be estimated by comparison with the solution of equations of type $\gamma$, see § 4.3 below.

where J and $j$ are new arbitrary constants; of these $j$ is small if $|\eta|$ is small at $t = 0$.[*] The motion is a combination of the following components:—

(1) A uniform rotation about the origin, represented by the term $e^{ip_1}$.

(2) A damping of the amplitude, represented by $(\sigma_0/\sigma)^{\frac{1}{2}}\,e^{-q_1}$.

(3) An oscillation of period determined by $p_2$ whose phase is continually changed by the factor $(j-q_2)$. The values of $\delta$ and $\phi$ are given by the equations

$$(4.031)\qquad \delta^2 = \tfrac{1}{2}J^2(\sigma_0/\sigma)\,e^{-2q_1}\,\{\cosh 2\,(j-q_2) - \cos 2p_2\},$$

$$(4.032)\qquad \phi = \phi_0 + p_1 + \operatorname{arc\,tan}\,\{\coth\,(j-q_2)\tan p_2\}.\dagger$$

So long as $(j-q_2)$ does not change sign, the average rate of increase of $\phi$ over any number of complete periods is $(p'_1 \pm p'_2)$.

Let $\alpha$ and $\beta$ be the successive maximum and minimum values of $\delta$ (assumed positive). In determining the values of $\alpha$, $\beta$, and the corresponding values of $t$, it is legitimate to neglect the changes of $q_1$, $q_2$, and $\sigma$, which are very small in a single period $p_2$. The maxima and minima are then given by putting $\cos 2p_2$ equal to $-1$ and $+1$ respectively in (4.031). Writing

$$(4.041)\qquad \alpha_1 = J\,(\sigma_0/\sigma)^{\frac{1}{2}}\,e^{-q_1}\cosh\,(j-q_2),$$

$$(4.042)\qquad \beta_1 = J\,(\sigma_0/\sigma)^{\frac{1}{2}}\,e^{-q_1}\sinh\,(j-q_2),$$

so that $\alpha_1$, $\beta_1$ are defined for all values of $t$, we have

$$(4.051)\qquad \alpha = \alpha_1\,(T_n),$$

for values of $T_n$ given by

$$(4.052)\qquad p_2\,(T_n) = \tfrac{1}{2}\,(2n+1)\,\pi,$$

$$(4.053)\qquad \beta = |\beta_1\,(T'_n)|,$$

for values of $T'_n$ given by

$$(4.054)\qquad p_2\,(T'_n) = n\pi.$$

An alternative expression for $\phi$ is then

$$(4.06)\qquad \phi = \phi_0 + p_1 + \operatorname{arc\,tan}\left(\frac{\alpha_1}{\beta_1}\tan p_2\right).$$

---

[*] The curves of $\delta$ against $t$ appear to have a minimum very near the muzzle of the gun in all rounds fired, but it will be seen that, in analysing the results, it is not necessary to assume any definite origin for $t$ or $\phi$.

[†] Here arc tan (A tan $x$) is determined in such a way that it changes continuously as $x$ increases indefinitely.

The curves of fig. 11 were calculated from formula (4.06), assuming $\alpha_1$ and $\beta_1$ constant, and $p_1$, $p_2$ proportional to $t$. They show the type of curve on which the observed values of $\phi$ may be expected to lie.

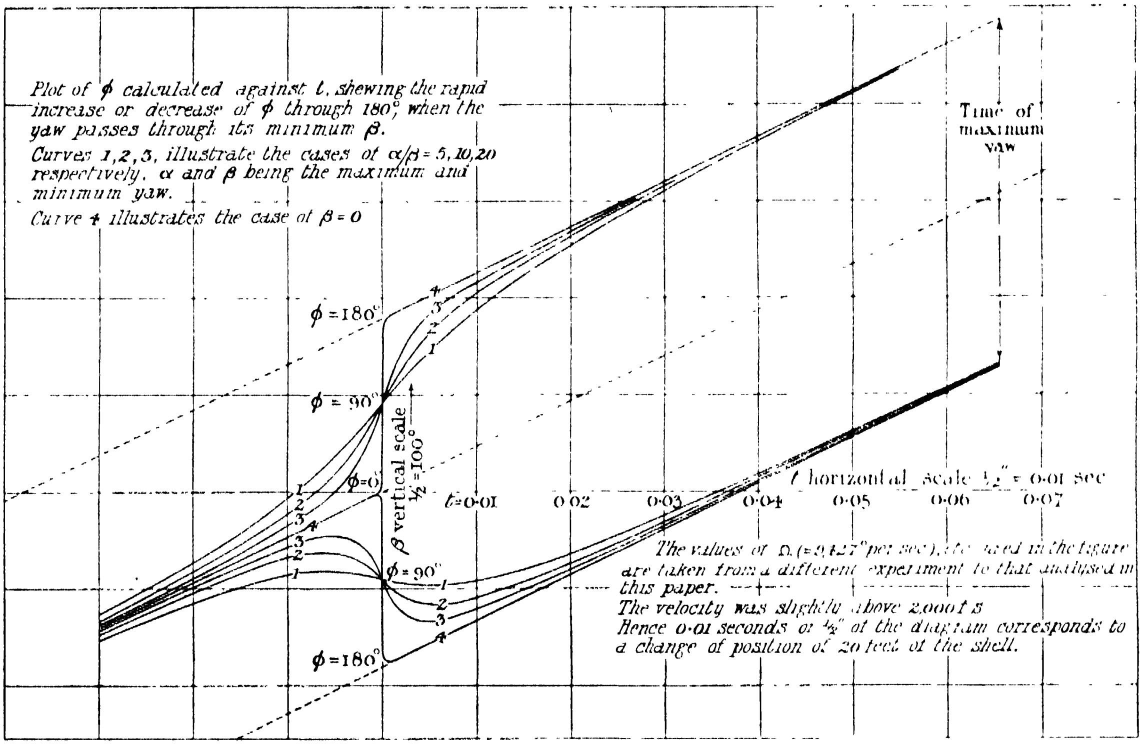


Fig. 11.

## § 4.1. *Analysis of the Experimental Results.*

It is now necessary to make use of these results to analyse the experiments. The analysis was carried out by graphical methods. The observed values of $\delta$ and $\phi$ were plotted on separate diagrams, examples of which are shown in fig. 12, against the abscissa $\Omega t$, $\Omega$ being determined from the muzzle velocity and the observed moments of inertia. The constant factor $\Omega$ was inserted to make the independent variable of zero dimensions; the values of the variable $\Omega t$, at given distances down the range, are also independent of small changes in the muzzle velocity. The observed values of $\delta$ are sufficient to give a good determination of curves showing the relation between $\delta$ and $\Omega t$, except in the neighbourhood of the minima $\beta$, where rapid changes of curvature occur when $\beta$ is small.* These curves give approximate values of the periods from minimum to minimum, and also the best determination available of the values and times of occurrence $T_n$ of the maxima $\alpha$. By drawing smooth (non-periodic) curves through the values of $\alpha$ we determine $\alpha_1$ as a function of $\Omega t$.

---

* When $\beta$ is small it may be convenient, in a preliminary plot, to change the sign of $\delta$ in each alternate period, so as to obtain smooth curves with $\delta$ passing through zero periodically.

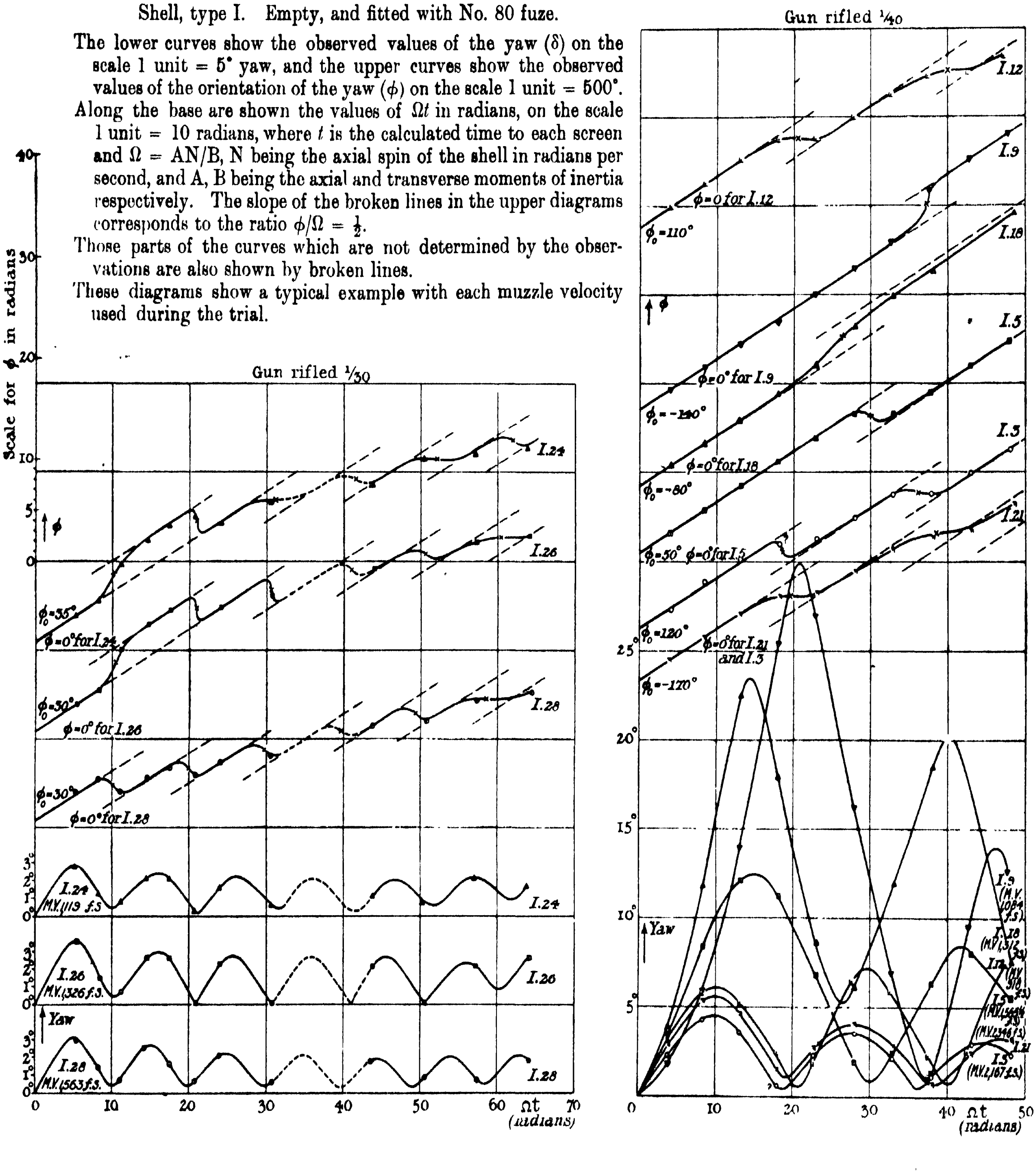
In drawing curves for $\phi$ against $\Omega t$ it is necessary to resolve the ambiguities of amount $2n\pi$ as follows:—Equation (4.06), or fig. 11, shows that $\phi$ increases by

Fig. 12A. Jump card trial, January and February, 1919.
3-in. 20-cwt. gun with 16-lb. H.E. shell, Mark IIB.
Shell, type I. Empty, and fitted with No. 80 fuze.

The lower curves show the observed values of the yaw ($\delta$) on the scale 1 unit = 5° yaw, and the upper curves show the observed values of the orientation of the yaw ($\phi$) on the scale 1 unit = 500°. Along the base are shown the values of $\Omega t$ in radians, on the scale 1 unit = 10 radians, where $t$ is the calculated time to each screen and $\Omega$ = AN/B, N being the axial spin of the shell in radians per second, and A, B being the axial and transverse moments of inertia respectively. The slope of the broken lines in the upper diagrams corresponds to the ratio $\phi/\Omega = \frac{1}{2}$.

Those parts of the curves which are not determined by the observations are also shown by broken lines.

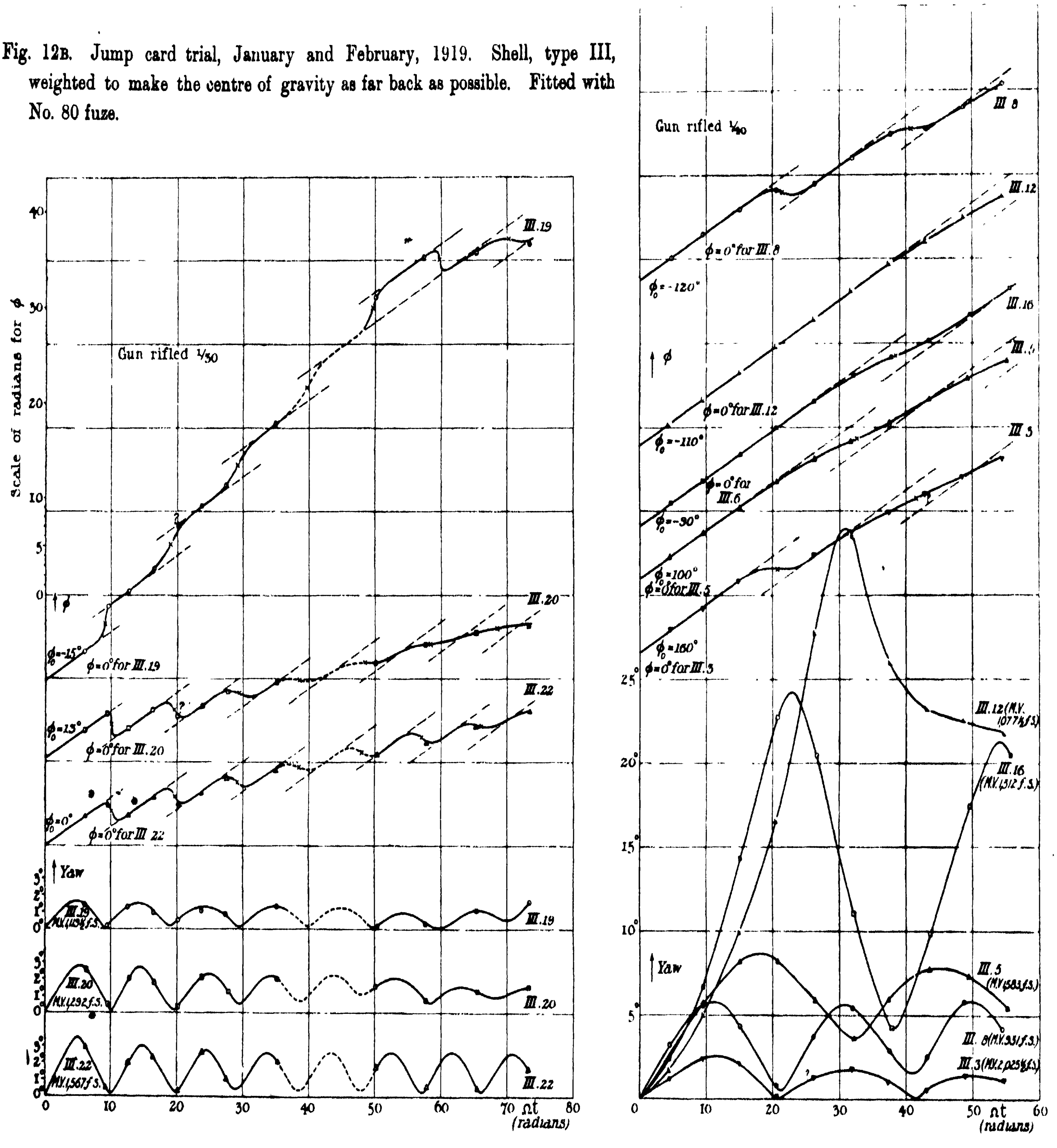
These diagrams show a typical example with each muzzle velocity used during the trial.



$\frac{1}{2}\Omega T \pm \pi$ in each period T, the increase of $\pm\pi$ occurring chiefly in the neighbourhood of the minimum, especially when $\beta_1/\alpha_1$ is small. Hence, by adding to the observed $\phi$

multiples of $\pi$, which are alternatively odd and even in successive periods of $\delta$, the points can be fitted roughly to a straight line of constant slope. All the points will lie fairly well on this straight line, except those in the immediate neighbourhood of the minima of $\delta$. By producing this straight line backwards we can determine the initial value of $\phi$. The slope of the straight line determines an independent value of $\Omega$, which is equal to the value deduced from the muzzle velocity if the slope is $0 \cdot 5$.

Fig. 12B. Jump card trial, January and February, 1919. Shell, type III, weighted to make the centre of gravity as far back as possible. Fitted with No. 80 fuze.



In practice the values of $\Omega$, obtained by the two methods, were in satisfactory agreement, except for the shells with centre of gravity forward, whose dynamical constants were considerably altered by the set back of the lead block on firing. For these shells the slope of the observed $\phi$-curve was taken as defining $\Omega$. The value of B after firing was deduced from this value, and the position of the centre of gravity was determined by equations (2.21) and (2.22).

The curve showing the true relation of $\phi$ to $\Omega t$ must pass through the true values of $\phi$, which differ from the observed values only by integral multiples of $2\pi$. It remains doubtful whether the value of $\phi$ increases or decreases by the amount $\pi$ radians in passing through a minimum of $\delta$, in addition to its steady increase at rate $\frac{1}{2}\Omega$. This question is settled by the divergence of points near the minimum of $\delta$ from the straight line of fig. 11; thus, if the points lie above the straight line in approaching a minimum, there will be an increase of amount $\pi$, and *vice versâ*. In this way a continuous curve may be drawn which is consistent with the equation (4.06). Specimens of the curves obtained in the analysis of the actual observations are shown in fig. 12. The portions of the curves in the neighbourhood of the maxima will then coincide approximately with a series of parallel straight lines at distances apart of $\pi$ radians. The method can only fail in one case when none of the points diverge appreciably from the straight line of fig. 11. This indicates that the value of the minimum $\beta$ is indistinguishable from zero, while the value of $\phi$ changes almost discontinuously by $\pm\pi$ at the time of the minimum. It is then immaterial whether the change is taken to be positive or negative.*

The observed values of $\phi$ in the neighbourhood of the minimum also yield information as to the value of $\beta_1/\alpha_1$ and the instant at which the minimum occurs. Let P be any observed point on a $\phi$-curve which diverges measurably from the nearest straight portion of the $\phi$-curve; lying above it by $\Delta$ degrees. Let $t_0$ be the time of occurrence of the nearest minimum, and $\delta p_2$ the change in $p_2$ between the minimum and P. Then, by (4.06),

$$\cot \Delta = \frac{\alpha_1 (t_0)}{\beta_1 (t_0)} \cot \delta p_2. \tag{4.101}$$

If, in this equation, $\Delta$, $t_0$, $\delta p_2$, and $\alpha_1 (t_0)$ are regarded as known, we can at once obtain a value of $\beta$. By adjusting the value of $t_0$ we attempt to reconcile the one or more values of $\beta$ obtained in this manner and also the value demanded by the $\delta$-curve. By combining all the available evidence in this manner, remembering that the $\delta$-curve is nearly symmetrical about a minimum, and the $\phi$-curve at the same time halfway between two straight portions, we can draw fairly precise final curves, obtaining values of $\beta_1(t)$ and the times of occurrence of the minima with some accuracy. Such curves are shown in fig. 12.

---

* The rapid changes or discontinuities in the values of $\phi$ and $\delta'$, which occur when $\delta$ is very small or zero, are due to the singularity which occurs at the origin of polar co-ordinates. The motion of the shell is, of course, in all cases continuous.

The following quantities have now been determined from the observations, viz.: the (assumed constant) value of $\Omega$, the times $T'_1$, $T'_2$, &c., of the occurrence of the minima of $\delta$ (those are more accurately determined than the times of the maxima), and the values of $\alpha_1(t)$ and $\beta_1(t)$ over the range of the experiments. These values are given in Table V.

4.11. *Derivation of the Various Force Components.*—It remains to derive the values of the various force components. By equations (4.054), (4.011)

$$(4.111) \qquad p_2(T'_n) - p_2(T'_{n-1}) = \pi, \qquad \int_{T'_{n-1}}^{T'_n} \Omega\sigma\, dt = 2\pi,$$

giving, as a sufficient approximation,

$$(4.112) \qquad \Omega\sigma = 2\pi/T \qquad (T = T'_n - T'_{n-1}),$$

$$(4.113) \qquad s = \frac{1}{1-(2\pi/\Omega T)^2},$$

$$(4.114) \qquad f_M\left(\frac{v}{a}\right) = \frac{A^2N^2}{4Bs\rho v^2 r^3},$$

where $\sigma$, $s$, and $v$ correspond to the time $\frac{1}{2}(T'_n + T'_{n-1})$. T is therefore the time between successive minima of $\delta$. The values of $s$ and $f_M$ obtained in this manner, or, in a similar way, taking an average over several periods, with the corresponding values of $\mu$ and $v/a$, are given in Table VI.,* and provide the data on which figs. 4 and 5 and Table I. were constructed.

By comparing the values of $f_M$ for shells of form A, with three different positions of the centre of gravity, the values of $f_L$ were deduced by the formulæ of § 1.13. This deduction was done graphically as shown in fig. 13. According to § 1.13 the relation between $f_M$ and $l$, the distance of the centre of gravity from the base of the shell, should be linear. Fig. 13 shows that all the observed points lie on straight lines within the limits of error of the observations. The slope of each line determines the value of $f_L$. The values of $f_L$ are shown plotted against $v/a$ in fig. 4.

---

* For the rounds fired from the gun rifled 1 in 30 the time of the first minimum near the muzzle is, in general, badly determined, and the first period is therefore omitted in determining a mean value for $s$. For the rounds fired from the gun rifled 1 in 40 the time of the first minimum can be determined with fair accuracy by extrapolation.

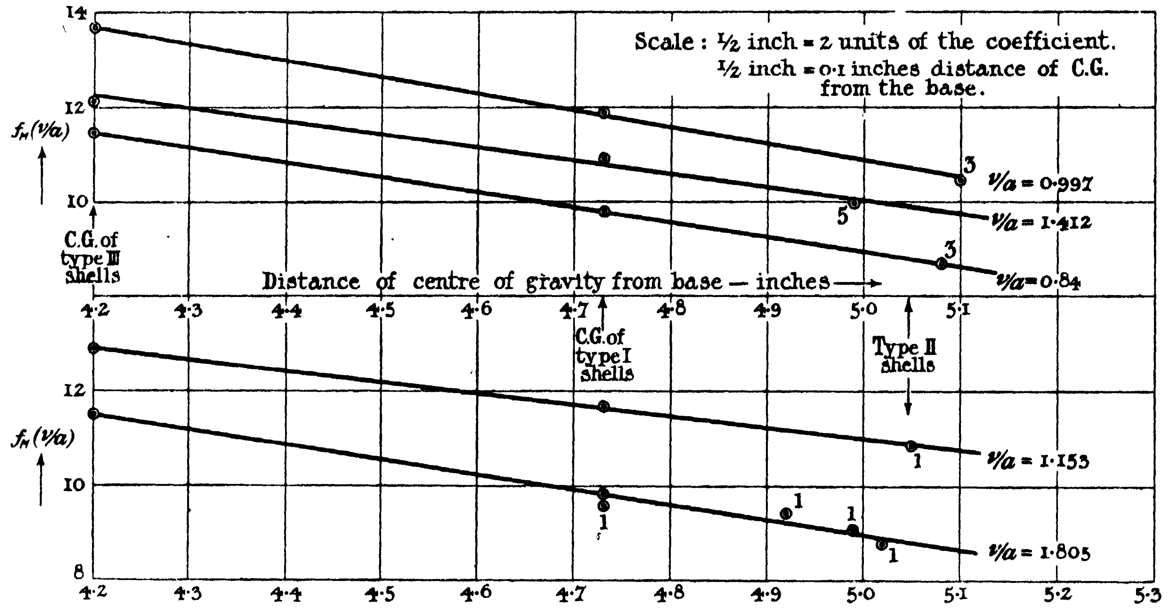


Fig. 13. The determination of the coefficient of the force acting normal to the shell.

The plotted points show the observed values of the couple coefficient plotted against the distance of the centre of gravity from the base.

The slopes of the lines drawn determine the coefficient of the normal force.

The numbers against the points for the Type II. shells give the number of observations whose mean is represented by the plotted point.

4.12. *The Damping Factors.*—It now only remains to derive as much information as possible as to the damping factors $\kappa$, $h$, and $\gamma$ from the observed values of $\alpha_1$ and $\beta_1$. The factor $\kappa$ is known in terms of the value of $f_L$, since, by § 3.1,

$$\kappa = \rho v r^2 f_L / m. \tag{4.121}$$

Squaring and subtracting equations (4.041) and (4.042), we obtain

$$\sigma_0 J^2 e^{-2q_1} = \sigma(\alpha_1^2 - \beta_1^2),$$

$$q_1 = -\tfrac{1}{2}\log\{\sigma(\alpha_1^2-\beta_1^2)\} + \text{const.},$$

$$h+\kappa = \frac{-1}{(t_2-t_1)}\Big[\log\{\sigma(\alpha_1^2-\beta_1^2)\}\Big]_{t_1}^{t_2}. \tag{4.122}$$

In this formula, as well as in those which follow, $\kappa$, $h$, and $\gamma$ may be treated as sensibly constant over the whole range of one experiment. On dividing (4.041) by (4.042), we obtain

$$\tanh(j-q_2) = \beta_1/\alpha_1. \tag{4.123}$$

Since $j$ and $q_2$ are both small over the range of the experiments, the formula becomes

$$h-\kappa+2\gamma = \frac{-1}{(t_2-t_1)}\left[\frac{\beta_1}{\alpha_1}\right]_{t_1}^{t_2}. \tag{4.124}$$

The three equations (4.121), (4.122), (4.124) for $\kappa$, $h$, and $2\gamma$ are in theory sufficient to determine their values completely. It may be noted again that $2\gamma$ is probably negligible and $h-\kappa+2\gamma$ is always positive, so that $q_2$ continually increases with the time, *and* $\beta_1/\alpha_1$ *continually decreases.* The constant $j$ is always very small, but may be positive or negative. If it is positive, $\beta_1$ is initially positive, giving the larger average rate of increase of $\phi$, which changes to the smaller rate of increase when $\beta_1$ becomes negative. If $j$ is negative, $\phi$ increases at the slower rate from the beginning. Exactly the opposite results would be obtained if $h-\kappa+2\gamma$ were negative. The values of $h+\kappa$ and $h-\kappa+2\gamma$, obtained in this manner, are given in Table VII.

In order to illustrate the actual path traced out by the axis of the shell, it is necessary to plot $\delta$ and $\phi$ as polar co-ordinates. This is done for three rounds in fig. 14. The resulting curves are roughly equivalent to the path of a point on the axis of the shell relative to the centre of gravity. They illustrate the decrease of $\alpha_1$, the algebraic decrease of $\beta_1$, and the tendency to change from quick to slow precession and to settle down to a *steady* slow precession.

The process described above was evolved gradually during the work of analysing the results, so that a number of observations were analysed before it was fully developed. It is probable that if the calculations were to be repeated *ab initio* a number of periods and minima of $\delta$ would be slightly altered, but it is unlikely that any serious systematic errors remain.

4.13. *Details of Tables V. to VII.*—The information contained in the General Table of Results, Table V., has been compiled by analysis of the original standard diagrams. As first constructed these were drawn with the time $t$ as abscissa and not $\Omega t$ as in fig. 12. It contains practically all the information of importance provided by the more stable shells. In the unstable cases, a number of which occurred during the trial (see for example fig. 12), a detailed study of the whole yaw curve is required which will not be undertaken in this paper.

Column 5 gives the values of the periods of the yaw curve in units of $\frac{1}{1000}$ second. The periods are read off from positions of the minima and sometimes of the maxima. They are entered to the nearest $\frac{1}{2000}$ second. They are in doubt by more than this quantity in many cases, but mainly in the case of the longer periods, in which small errors are of less importance.

Column 6 gives the values of the maxima of the yaw in degrees and decimals to one place of decimals. These values are read straight from the curves and represent roughly the accuracy to which the maxima are in most cases determined by the observations.

Column 7 gives two entries. The first is the value of $\Omega T$ for each round, where T is the mean value of the observed period and $\Omega$ corresponds to the *observed* value of the steady rate of increase of $\phi$ from column 4.

The second entry in column 7 is the velocity of the shell at the middle point of the range of periods whose mean value T is used to determine $\Omega T$. The stability factor determined by $\Omega T$ is taken to correspond to this velocity Finally, in column 8, the values of $\beta_1(t)$ are given with their proper sign as determined incidentally in the determination of their times of occurrence (§ 4.11).

The effect of the cards on the observed value of the period and on $s$, is ignored in Tables V. and VI. The results obtained here are corrected for this effect, as far as possible, before use in Table I. The information given in Table VI. is deduced directly from Table V. by the equations of § 4.11. In certain cases where the yaw was large it was checked by use of the chart of § 3.7.

The total percentage spread of the values of $s$ (or $\mu$) in the group is in most cases satisfactorily small. The value of 6·7 per cent. for the high velocity group of type I. shells is probably partly due to the fact that the fuzes of shells 1 to 4 were slightly damaged before firing in forcing the shells into the cartridge cases.

At a velocity of 1580 f.s. results were obtained with guns of both twists of rifling. The couple deduced from the results for the gun rifled one turn in 40 calibres is, in the cases of shells of types I. and III., slightly smaller than that deduced from the other gun. This is to be expected as the stability in this case is nearly critical and the maxima are rather large (one maximum is as much as 13 degrees for a type I. shell). The solution of § 3.6 can hardly be expected to apply. The next term in the expression for $\mu$ of the form $\mu_1 \sin^3 \delta$ may be expected to be becoming appreciable here; apparently its sign is such that it will tend to diminish the observed value of $\mu$, in agreement with wind channel observations (fig. 2). For the shells of type II. the maxima of the yaw are small in both guns and the results are in agreement.

No perceptible dependence of $s$ on the maximum yaw among the rounds of any one group has been detected in these tables.

The agreement between the results for the two guns at this velocity, and between rounds with different maxima of the yaw, is therefore a satisfactory confirmation of the theory.

The values of $h+\kappa$ and $h-\kappa+2\gamma$, deduced from the observations as explained in § 4.12, are given in Table VII. Of these, the former is more reliable as it does not depend on $\beta_1(t)$ which is difficult to determine. The actual values vary considerably from round to round, and only mean values for each group are shown. The results are therefore very rough, but they indicate qualitatively the nature of the damping, which may also be studied in figs. 12 and 14. For example, in fig. 14C, the motion starts with $\beta_1(t)$ positive, so that the loop encloses the origin, O, or point of zero yaw. But since $h-\kappa+2\gamma>0$, $\beta_1(t)$ diminishes and has become negative by the second minimum, the loop failing to reach O. As $\beta_1(t)$ diminishes further, the loop shrinks to a cusp at the fourth minimum, and the motion soon becomes indistinguishable from a precession at the slower rate. In the meantime, the maximum yaw $\alpha_1(t)$ decreases steadily.

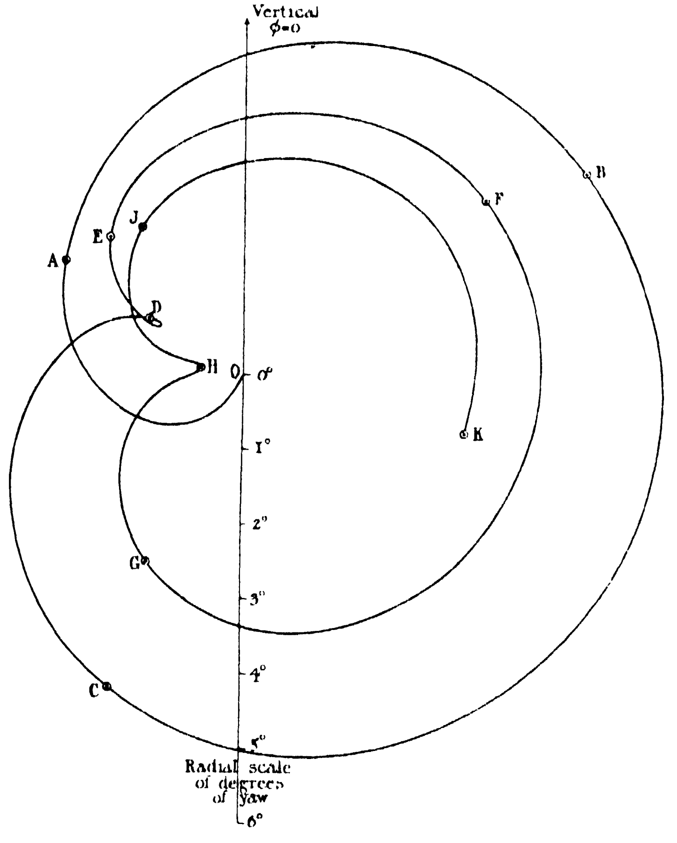


Fig. 14A. Path of nose of shell. Round I.21.

Path described, relative to the centre of gravity, by a point on the axis of the shell in front of the centre of gravity, shown on an enlarged scale.

The total time taken from O to K is 0·2572 second. On the scale used, 1 cm. distance from O represents 1° yaw (very nearly), and corresponds to a linear displacement of 0·118 inch for the nose of the shell from the line of motion of the centre of gravity.

The numerical results for the damping must be affected to some degree by the impacts on the cards, but the available data are not good enough for corrections to be worth making. There is, moreover, the curious phenomenon of an increasing maximum yaw shown by the rounds at 900 f.s. to be accounted for.

The value of $\kappa$ is known from equation (4.121) and the values of $f_L$ in Table I., so that the damping results determine $h$ and $h+2\gamma$ or, more accurately, $h+2\gamma-\Gamma$ (§ 3·62). It at once appears that $2\gamma-\Gamma$ is negative and of much the same order as $h$. This is somewhat unexpected. Of course $\Gamma$ (or $-N'/N$) is positive, but it is hardly likely that its numerical value is much larger than 0·03. It is natural to expect

$\gamma$ to be small and positive,* which does not fit in with the observations. Further experiments would be needed to throw light on all these points.

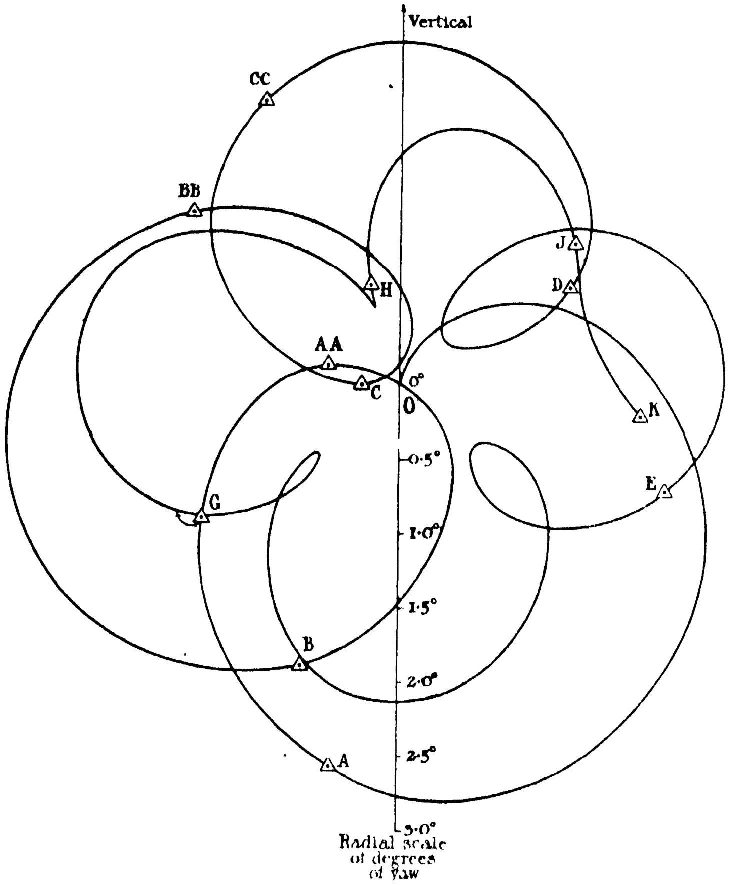


Fig. 14B. Path of nose of shell. Round III.20.

Path described, relative to the centre of gravity, by a point on the axis of the shell in front of the centre of gravity, shown on an enlarged scale.

The total time taken from O to K is 0·3647 second. On the scale used, 2 cm. distance from O represents 1° yaw (very nearly), and corresponds to a linear displacement of 0·128 inch for the nose of the shell from the line of motion of the centre of gravity.

In the fourth column what appears to be the most probable value of $h$ is given; the values of $f_H$ in Table II. are based on these figures and obtained by the equation (see §§ 3.5 and 1.12)

$$f_H = \frac{hB}{\rho v r^4}.$$

---

* The coefficient $\gamma$ comes from the swerving couple J (§ 1.12). This couple will only arise if the swerving force K does not act through the centre of gravity. Since the air pressures are greater near the nose than near the base, we may expect K to act *in front* of the centre of gravity. By analogy with the connection between the direction of rotation and the direction of the resulting swerve on a golf or tennis ball at low velocities, we may expect K to act along the axis of M reversed in fig. 9, for a right-handed twist of rifling. This would result in a positive value for $\gamma$.

The figures in Table VII. were obtained from the sufficiently stable rounds fired from either gun. In the one comparative pair of groups available the results for the two different stability factors and values of $\Omega$ were in agreement.

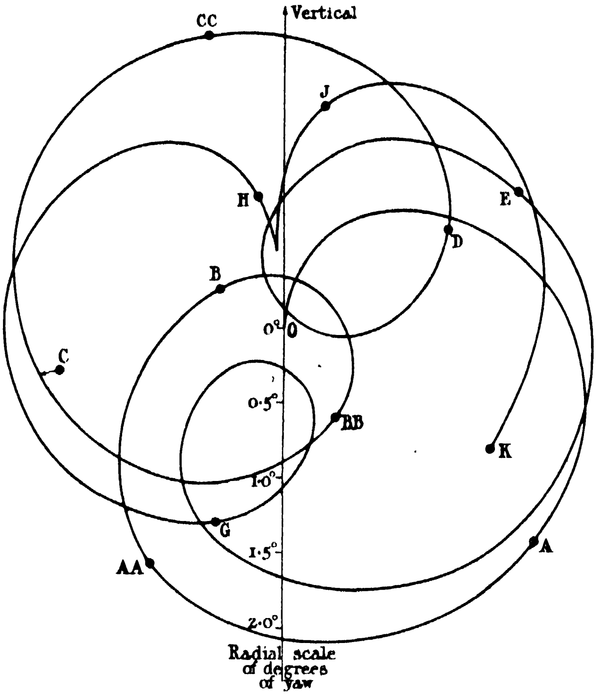


Fig. 14c. Path of nose of shell. Round IV.15.

Path described, relative to the centre of gravity, by a point on the axis of the shell in front of the centre of gravity, shown on an enlarged scale.

The total time taken from O to K is 0·5502 second. On the scale used, 2 cm. distance from O represents 1° yaw (very nearly), and corresponds to a linear displacement of 0·143 inch for the nose of the shell from the line of motion of the centre of gravity.

Note that the first loop encloses O, corresponding to the "stepped up" motion in $\phi$. Subsequent loops do not, as the motion in $\phi$ has changed to the "stepped down" motion. (See fig. 12.)

### § 4.2. *Determination of the Motion of the Shell in Space.*

We now proceed to make use of the results of the experiments to determine the true path of the centre of gravity of a shell projected in a given manner. The solution of the equations of type $\alpha$ is sufficient for this purpose so long as the yaw does not exceed 0·1 radian; the values of $f_L$, $f_M$, $f_H$, &c., which we have obtained, are sufficient to determine the motion completely in this case. Assuming that the maximum yaw due to the initial disturbances is less than 0·1 radian in the first period, it will remain so throughout the trajectory; the yaw arising from the particular integral will not exceed 0·1 radian until the velocity has fallen considerably below 700 feet per second. Hence, when the yaw exceeds 0·1 radian, the wind channel values for the various force components as shown in fig. 2 can be used; it will, however, be necessary to abandon the above method of solution and proceed by the step-by-step integration of the equations of type $\beta$.

Throughout the following work all numerical results will be based on a set of plane trajectories of a 16-lb. shell, of external form A, fired at a muzzle velocity of 2000 f.s., calculated by the ordinary ballistic methods.* The various elements of the trajectories at elevations of 30 degrees and 50 degrees and a list of constants for the service shell, to which the calculations apply, are given in Table VIIIA.

From the value of $\zeta$ for the general solution, as given in § 3.65, we can deduce the true path of the centre of gravity in terms of the tabulated elements of the plane trajectory. Let $(X_1, Y_1, 0)$ be the co-ordinates of the shell in the plane trajectory at time $t$, and $(X, Y, Z)$ the corresponding co-ordinates in the true (twisted) trajectory. The direction cosines of the tangents to the two trajectories are $(X'_1/v_1, Y'_1/v_1, 0)$, or $(\cos\theta_1, \sin\theta_1, 0)$ and $X'/v$, $Y'/v$, $Z'/v$, so that, to the usual order of approximation,

$$(4.201)\qquad \begin{aligned} c\zeta &= (Y'-Y'_1)(\cos\theta_1)/v_1 - (X'-X'_1)(\sin\theta_1)/v_1 + iZ'/v_1 \\ &= (H'+iZ')/v_1, \end{aligned}$$

say, while the condition $v = v_1$ gives

$$(4.202)\qquad (X'-X'_1)\cos\theta_1 + (Y'-Y'_1)\sin\theta_1 = 0.$$

It is convenient to separate the parts of the solution arising from the complementary function and the particular integral. To determine the latter, we use equations (3.632), (3.633), and (4.201), obtaining

$$(4.203)\qquad \begin{aligned} \frac{Z'}{v_1} &= c\int_0^t \frac{-4s\kappa\theta_1'\,dt}{c\Omega} \\ &= c\psi, \end{aligned}$$

say, neglecting the terms $i\theta''_1/\Omega$ in $\Phi$ (see § 3.20). This equation defines $\psi$. Therefore

$$(4.204)\qquad Z = \int_0^t \psi v_1 \cos\theta_1\,dt,$$

where $\psi$ may be written (since $-\theta'_1/c = g/v_1$)

$$\psi = \int_0^t \frac{g\kappa AN}{\mu v_1}\,dt = \frac{Ag}{mr}\int_0^t \frac{N f_L(v/a)}{f_M(v/a)}\frac{dt}{v^2}.$$

To the same approximation $(X'-X'_1)/v_1$ and $(Y'-Y'_1)/v_1$ are $O(1/\Omega^2)$, so that $(X_1-X)$ and $(Y_1-Y)$ are small compared to $Z$, so long as the approximations hold. The above result is identical in form with the "classical" formula of **Mayevski**, freed from the unnecessary restriction that $f_L/f_M$ and N should be constants.* We have thus justified the use of the plane trajectory as an approximation to the true motion. The leading terms in $(X-X_1)$ and $(Y-Y_1)$ can be calculated if required.

---

* Trajectories were calculated with the ballistic coefficient 1·75.

The effect on the motion of a change in initial conditions is obtained from the complementary function. Equation (3.655) gives the value of $\zeta$ corresponding to the general initial conditions $\zeta_0 = 0$, $\eta_0 = a$, $\eta'_0 = b\Omega$, where $a$ and $b$ may be complex. Substituting in equation (4.201) the part of $\zeta$ arising from the complementary function, it appears that $H+iZ$ is made up of three parts :—

(*a*) A periodic term

$$H_1+iZ_1 = -\frac{4\kappa v_1}{\Omega^2}\left(\frac{K_1\eta_1}{(1+\sigma)^2}+\frac{K_2\eta_2}{(1-\sigma)^2}\right).$$

(*b*) A term

$$H_2+iZ_2 = -\{K_1(\zeta_1)_0+K_2(\zeta_2)_0\}\int_0^t cv_1\zeta_3\,dt,$$

which is the effect of a variation in the direction of projection, as mentioned in § 3.64.

(*c*) A constant term $H_3+iZ_3$ equal to the initial value of $H_1+iZ_1$ with its sign changed.

4.21. *Numerical Results as to Motion of Centre of Gravity.*—The only data as to the forces on the shell required for the calculation of the drift are the value of $f_L/f_M$ as a function of $v/a$. This is derived from the results of the jump card experiments for $v/a > 0\cdot7$, and from wind channel experiments for $v/a < 0\cdot7$, and is shown plotted in fig. 15.

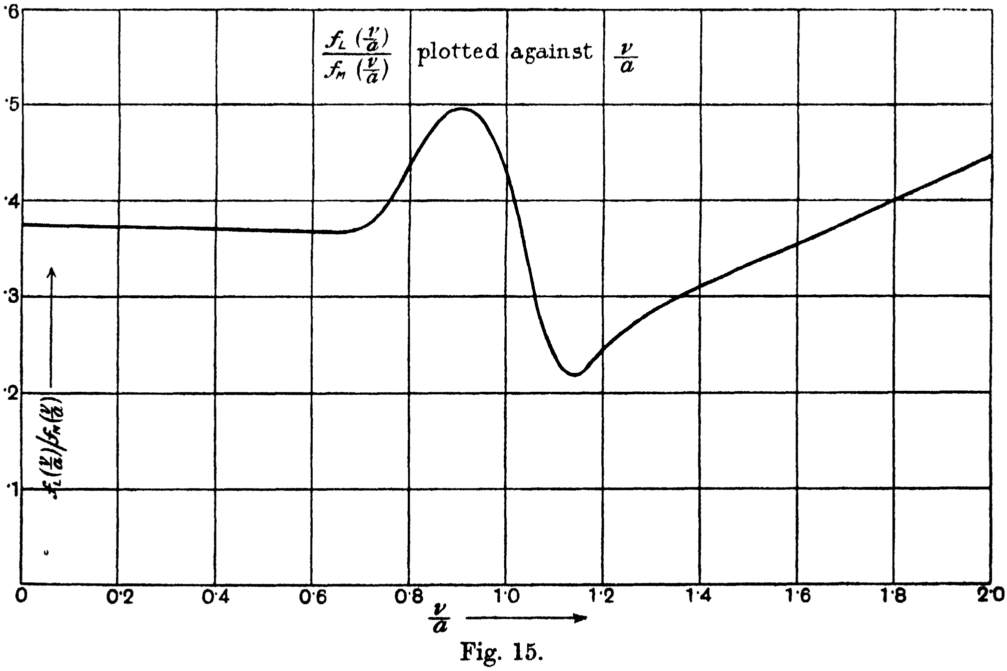


Fig. 15.

---

* PRESCOTT obtains a solution of the equations of motion in the form of a series of which the first term is also equivalent to MAYEVSKI'S formula. (See Introduction, p. 296.)

As the value of the couple $\Gamma$ is only known to be small, it is necessary to assume that N is constant. The principal steps in the calculation of the drift Z, by means of (4.203) and (4.204), for the trajectories at 30 degrees and 50 degrees, are given in Table VIIIb. for the gun rifled 1 turn in 30 diameters of the bore. For a different rifling the drift (N constant) is proportional to N.

It is only necessary to estimate roughly the effect of the complementary function on the motion of the shell since the total effect is always fairly small. The periodic term $H_1+iZ_1$ is obviously smaller than $4\kappa v_1\alpha_1/\Omega^2(1-\sigma)^2$ in absolute value, where $\alpha_1$ is defined as in § 4.0, equation (4.041). The initial value of the coefficient of $\alpha_1$ is $1\cdot25$ feet for the gun rifled 1 turn in 30 diameters of the bore; both $\alpha_1$ and its coefficient diminish rapidly. Taking $\alpha_1 = 0\cdot1$ radian as an extreme case,

$$|H_1+iZ_1| < 1\cdot5 \text{ inches.}$$

The actual value in practice is probably always $< 0\cdot5$ inch, which is small. It explains why no evidence of helical motion was obtained in the jump card experiments. The constant value of $|H_3+iZ_3|$ is equal to the initial value of $|H_1+iZ_1|$ and is also negligible. There remains only the term $H_2+iZ_2$. This is equivalent to an angular deviation of $|K_1(c\xi_1)_0+K_2(c\xi_2)_0|$, which is less than $2\kappa\alpha_1/\Omega(1-\sigma)$. The coefficient of $\alpha_1$ for the gun rifled 1 turn in 30 diameters of the bore is $1\cdot8\times10^{-2}$, so that for a value of $\alpha_1$ of 0·1 radian the angular deviation is of the order 0° 6′. This is of the same order of magnitude as the angular jump likely to be due to changes of form and position in the gun and mounting under firing stresses. When it varies from round to round in magnitude and direction, it will account for an *irregularity* of the corresponding amount in the observed positions of the shells at any time. When, as may sometimes be the case, it remains fairly constant from round to round, it will cause systematic errors in the position of the shell at any time. It is probable that anomalous values of the drift, sometimes observed at short times, are due to this cause. Practical results, however, more often fully justify the use of the particular integral alone to give a mean value of the drift when the initial disturbance is only known to be small.

The results of the above calculations of drift will now be compared with observations of the Z co-ordinates of the bursts of shells, fired at Portsmouth, at corresponding elevations, in February and April, 1918. For this purpose use is made of the azimuth of the shell burst Z/X; the quantity $A = Z/Xt$ is tabulated, since its value varies slowly along the trajectory (Table IX.). The agreement between observation and calculation is as good as could be expected, in view of the uncertainty in the wind effects, and provides important evidence as to the correctness of the whole theory.

4.22. *The Damping of the Angular Oscillations and the Effect on the Head Resistance.*—We have now obtained the complete motion of the centre of gravity of the shell by use of the equations of type $\alpha$ for the two standard trajectories; we have, in so doing, assumed that the velocity of the shell is the same in the plane and true trajectories; we must now examine more closely the possible effect on the drag of the angular oscillations and their rate of damping, by means of the values of $h$ and $\kappa$ obtained above. From equations (4.031) and (4.032) it appears that for sufficiently large values of $t$, $\delta$ and $\phi$ are given approximately by the equations

$$\delta = \tfrac{1}{2}\mathrm{J}\,(\sigma_0/\sigma)^{\frac{1}{2}}\,e^{-(q_1-q_2)},$$
$$\phi = \phi_0 + (p_1 - p_2),$$

so that the shell settles down to a steady precession with the slower precessional angular velocity, the yaw gradually diminishing in proportion to the factor $(\sigma_0/\sigma)^{\frac{1}{2}}\,e^{-(q_1-q_2)}$. This quantity is tabulated in column 9, Table VIIIB., on the assumption that $\gamma = 0$ and $h = 3\kappa$. The damping is actually more rapid than is indicated by this approximation.

The question of the rate of damping of the initial oscillations of a shell is of importance on account of its effect on the drag R, for the effect, though it may be small, will be cumulative, since it tends always to increase R. If it is assumed that the effect on R is given by*

(4.221)
$$\mathrm{R} = \mathrm{R}_0\,(1 + k\delta^2),$$

where $\mathrm{R}_0$ is a function of $v$ and $k$ is a constant, it is possible to obtain an approximate formula for the total change in velocity produced on the assumption that the time taken to damp out the oscillations is relatively small. We have

(4.222)
$$-\Delta v = \frac{1}{m}\int_0^\infty (\mathrm{R}-\mathrm{R}_0)\,dt,$$
$$= \frac{k\mathrm{R}_0}{m}\int_0^\infty \delta^2\,dt.$$

Using (4·031), and integrating on the assumption that $\sigma$ is constant, $q_1$, $q_2$, and $p$ proportional to $t$, and $j$ zero, we obtain

(4.223)
$$-\Delta v = \frac{k\mathrm{J}^2\mathrm{R}_0}{2m}\int_0^\infty e^{-2q_1}\,(\cosh 2q_2 - \cos 2p_2)\,dt,$$
$$= \frac{k\mathrm{J}^2\mathrm{R}_0\sigma_0{}^2\,(h_0+\kappa_0)}{2m\,\{\sigma_0{}^2\,(h_0+\kappa_0)^2 - (h_0-\kappa_0+2\gamma_0)^2\}}\,.$$

At present we have no information as to the value of $k$ except at low velocities, while J varies from round to round so that no numerical results can be given. It seems likely that this is a cause of irregularities in range in practice of first class importance. The yaw arising from the particular integral will also tend to increase the resistance, but the effect is of less importance in a low angle trajectory.

---

* By symmetry, there can be no odd powers of $\delta$ in R.

4.23. *The Exact Motion in a High Angle Trajectory.*—It will be shown in the next section that, for a trajectory of much higher elevation than 50 degrees, the approximations for the particular integral break down, and the equations of type $\alpha$, are not applicable to the later stages of the trajectory when the velocity has fallen much below 500 f.s. These later stages occur after the initial oscillations have been damped out, and are suitable for the use of equations of type $\beta$. These equations can be integrated step-by-step on the basis of the wind channel values of R, L, and M (fig. 2), which apply to velocities up to 700 f.s. The process is analogous to the usual method of calculating a plane trajectory, but rather more laborious, and has been carried out in one case only for a 3-inch 12½-lb. shell fired at 70 degrees with a muzzle velocity of 2450 f.s. At 40 seconds the yaw has reached the large value of 60 degrees and is still increasing. This is partly due to the large initial value of the stability factor (about 4·0) indicating that the spin is unnecessarily large for this shell. The results of comparing the drift with observation were again fairly satisfactory in this case; but details of these results cannot be given here.

### § 4.3. *Estimate of the Errors in the Various Solutions.*

In the development of the various solutions of the equations of motion in Part III., it was found necessary to neglect certain terms. We shall now proceed to examine these terms in succession, and to determine, as far as possible, their numerical values, using the values of the various force components obtained from our experiments. By so doing we shall justify the use of the solutions by showing that the terms neglected are all very small over the range covered by the jump card experiments. In the applications to the later parts of a trajectory, the solutions break down in certain cases, and an examination of the error terms enables us to define the circumstances under which the solutions are valid. We proceed to examine the various terms. It is necessary as a rule to distinguish the terms neglected in obtaining the complementary function from the terms neglected in obtaining the particular integral.

In the complementary function, $m$, $n$, $y$, $z$ are periodic functions of the time with periods comparable with $\Omega$. For the solution to be applicable we have to assume that $\delta$ is always small (say $\delta < 0 \cdot 1$ radian). Then $m$, $n$, $y$, $z$ are all small quantities comparable with $\delta$, and $m'/\Omega$, $m''/\Omega^2$, &c., are also comparable with $\delta$, while $(1-l)$, $l'/\Omega$, $l''/\Omega^2$, &c., are of the order of $\delta^2$. In neglecting terms independent of $\theta'_1$ from the equations (3.202), (3.203), we are guided by the condition that all terms neglected should be of the order of $\delta^2$ compared with those retained. As regards the terms containing $\theta'_1$ or $\theta''_1$, it appears that the maximum value of $\theta'_1/\Omega$ in the 50 degrees trajectory (for rifling 1 in 30) is $30 \times 10^{-5}$, its initial value being $5 \times 10^{-5}$. Hence all terms such as $nm'\theta'_1$, $n\theta'^2_1$ are completely negligible in obtaining the complementary function.

If all terms in $\theta'_1$, $\theta''_1$ are removed from equations (3.202), (3.203), they become equivalent to the equations of type $\gamma$; the errors in neglecting further terms may, therefore, be determined by comparing the solutions of equations (3.204), (3.205), &c., of type $\alpha$ with those of equations (3.404), (3.405), of type $\gamma$ (assuming $\mu$ constant in both cases). Equations (3.7041), (3.7043), (3.7044), (3.705) can be made to give the following approximation to the true value of $s$ in terms of T and $\alpha$:—

$$s = \frac{1}{1-4\pi^2/\Omega^2\mathrm{T}^2}\left\{1-\tfrac{1}{8}(2s+1)\alpha^2\right\}.$$

This is valid so long as (1) $\alpha$ is so small that $\alpha^4$ may be neglected, and (2) $s-1$ is positive and large compared with $\alpha^2$. Comparing this with the corresponding first approximation (4.113), we obtain the error in the value of $s$ due to the neglect of the terms in $(1-l)$, $l''$, &c., in equations (3.202), (3.203). The relative value of the error is given in the following table:—

| $\alpha$ | $s = 1{\cdot}1$. | $s = 2$. | $s = 3$. |
|---|---|---|---|
| 10° | 0·012 | 0·019 | 0·027 |
| 5° | 0·0030 | 0·0048 | 0·0066 |
| 2·5° | 0·0007 | 0·0012 | 0·0016 |

In analysing the jump card trial, whenever the error from this cause is appreciable, the results have been corrected by determining the values of $s$ from the chart described in § 3.7.

It appears also from the solution of the equations of type $\gamma$ that when $s \leq 1$ the initial angular motion is still periodic, but no longer of the nature of a small oscillation, since the period is a function of the amplitude and tends to infinity as the initial disturbance tends to zero.

In using equations (3.202), (3.203) to obtain the particular integral, the order of magnitude of the various terms is different. The term $\mathrm{AN}l\theta'_1$ is now the most important, while $n$ is $\mathrm{O}(1/\Omega)$ and $m$ is $\mathrm{O}(1/\Omega^2)$ with the notation of § 3.6. Most of the terms neglected are then $\mathrm{O}(1/\Omega^4)$ compared to the principal term, and completely insignificant, but $\mathrm{B}nl\theta'^2_1$ is $\mathrm{O}(1/\Omega^2)$ and would affect the third term in the expansion for $\eta$. Its effect however is completely negligible.

4.31. *The Equations of Motion of the Centre of Gravity.*—These equations may be treated in a similar manner. In obtaining the complementary function, $y$ and $z$ are small compared to $m$ and $n$ (see equations (3.624), (3.625)), $\kappa/\Omega$ being initially less than 0·01. As regards the differential equation for $u$ (3.2141), the effect of neglecting the terms arising from the variation of R with $\delta$ has been discussed in § 4.22; no numerical data are available; the effect is theoretically second order. The term in $1-x$ is obviously negligible. Omitting these terms, the equation can be reduced to the form

$$u'+\alpha u = -gy\cos\theta_1,$$

where

$$\alpha = \frac{1}{m^*}\frac{\partial}{\partial v_1}\{R(v_1+\theta u,\ \theta\delta)\} \qquad (0<\theta<1).$$

As a rough approximation we may assume that $R = kv^2$, so that $\alpha = -2v'_1/v_1$. We then find that

$$\frac{d}{dt}\frac{u}{v_1^2} = -\frac{yg\cos\theta}{v_1^2}.$$

In the case of the complementary function, $y$ consists of a constant term less than $2\times10^{-3}$, and periodic terms whose period is of order $1/\Omega$.

The former makes a contribution to $u/v_1$ which is still less than $10^{-3}$ after 20 seconds. The latter contributes† a term of order $yg/v_1\Omega$ which is always less than $3\times10^{-6}$.

In the case of the particular integral $y$ is $O(1/\Omega^2)$, see § 4.2. Hence in all cases we are justified in putting $u=0$, $v=v_1$, so long as the equations of type $\alpha$ hold at all, with the proviso that this conclusion may be at fault if the $k$ of § 4.22 is numerically large.

In reducing equations (3.212) and (3.213) we put $x=1$, $\cos\delta=1$. This amounts to neglecting $1-x$, $1-\cos\delta$ compared to 1, and is obviously justifiable. We omit altogether from (3.212) the terms $x\theta'_1+(g/v)\cos\theta_1$, or $-g\cos\theta_1\,(x/v_1-1/v)$. This term is excessively small, but could be retained, if desired. Finally we omit the terms in $y\cos\theta_1$, justifying the omission by the arguments used above for the same term in the equation for $u$.

§ 4.32. *Errors in the Solution for the Complementary Function.*—The second term $R_1$ in the expansion of R in equation (3.6233) will be taken as representing the principal part of the error in the standard solution for the complementary function arising at this stage. Its value is

$$R_1 = -\tfrac{1}{4}M^{-\frac{3}{4}}\frac{d^2}{dt^2}(M^{-\frac{1}{4}}),$$

where M has the value appropriate to (3.6231). For simplicity in estimating errors we may take only the leading term in M so that here

$$M = \tfrac{1}{4}\Omega^2\sigma^2.$$

The values of $s$ determined from the jump card trial and the data of the 50 degrees plane trajectory are tabulated in column 2 of Table VIIIB. The value of $\sigma$ can be deduced and its first and second differential coefficients obtained from a difference formula. In this way we find that initially

---

† This contribution is of the form $\int_0^t f(t)\,e^{i\Omega t}\,dt$, which is of the order $(1/\Omega)\times(\text{maximum of } f(t))$ under suitable restrictions on $f(t)$, which are satisfied here.

$$R_1/R_0 = 0\cdot000011,$$

$$R_1/R_0 = 0\cdot0031,$$

for the guns rifled 1 turn in 30 diameters and 40 diameters respectively. Moreover, the value of $R_1/R_0$ diminishes along the trajectory. The neglect of this term is therefore justified, provided $s > 1\cdot1$, and the total error in the solution will probably be of the same numerical order.

The contribution of $\zeta$ (see 3.6235) to the coefficient of $\eta$ in equation (3.613) is

$$-\frac{2\,\kappa c'}{c\,(1 \pm \sigma)},$$

the principal term in this coefficient being $-\Omega^2/4s$. The relative value of the error in omitting this term is therefore

$$\left|\frac{8sc'\kappa}{c\Omega^2\,(1 \pm \sigma)}\right|,$$

which is less than

$$\left|\frac{4s\kappa}{\Omega}\,\frac{4s\theta'_1}{\Omega}\tan\theta_1\right|.$$

The values of these three factors can be obtained from Table VIII. The maximum value of this ratio for the 50 degrees trajectory is $\frac{1}{250}\tan\theta_1$. This is negligible.

In evaluating M to obtain equation (3.6234), terms such as $\kappa^2/\Omega^2$, $\kappa h/\Omega^2$, $\kappa'/\Omega^2$ are neglected. It is unnecessary to evaluate such terms in detail, since it is known that $\kappa'/\Omega$ and $h/\Omega$ are less than $0\cdot02$ in all cases. It would, however, be easy to write down equation (3.6234) with such terms included.

4.33. *Errors in the Particular Integral.*—The errors of the expression for the motion of the centre of gravity of the shell, given in (3.632) and (3.633), may be obtained from the expansion of the particular integral in powers of $1/\Omega$. The ratio $i\theta''_1/\Omega\theta'_1$ of the two terms in $\Phi$, § 3.2, can be worked out from the data of the plane trajectory. Its initial and greatest value for the gun rifled 1 turn in 30 diameters is $(0\cdot0008)$, so that the second term is entirely negligible in comparison with the first. Writing therefore $\Phi_1 = \theta'_1$, it appears that the terms of order $1/\Omega^2$ in (3.632) are real and so do not affect the drift. The next term is (with $\gamma = 0$)

$$\frac{4s}{(i\Omega)^3}\left\{-\eta''_1 + \frac{d}{dt}\,(4s\eta'_1) + \eta'_1\,(8s\kappa - h - \kappa) + \eta_1\left[\frac{d}{dt}(4s\kappa) + 4s\kappa^2 - h\kappa - \kappa'\right]\right\},$$

where $\eta_1\ (= 4s\theta'_1)$ is the coefficient of $1/i\Omega$ in the first term in the expansion of $\bar{\eta}$.

There are also a number of other terms involving $c'$, $c''$ and $c'^2$. The terms in $c'$ are very small initially and vanish at the vertex, so that they are never likely to become important. The other terms in $c''$ are certainly very small provided that $s$ is of order unity. Since $s$ varies roughly inversely as the square of the velocity (*i.e.*, $f_M$ constant), the magnitude of the terms containing $s$ rises very rapidly in the later stages of the trajectory when $v$ becomes small. The first term in $\bar{\eta}$, $-4is\theta'_1/\Omega$, is given numerically in Table VIII., where it appears how rapidly it increases as the velocity falls. The values of the second term as given in equation (3.632) are also given (Table VIIIB., column 8). It appears that the ratio of the second term to the first term is always small so long as the first term is small. This term represents the effect of the particular integral in altering the co-ordinates in the plane of fire. The third term as given above is more difficult to evaluate, and only a rough estimate has been made of its value at two points on the 50 degrees trajectory. The results are:—

| Seconds. | Third term/first term. | Third term. |
|---|---|---|
| $t = 0$ | $-2{\cdot}02 \times 10^{-8}$ | $-8{\cdot}5 \times 10^{-7}$ |
| $t = 20$ | $-1{\cdot}94 \times 10^{-2}$ | $-7{\cdot}2 \times 10^{-4}$ |

The value of the drift as estimated by the first term is therefore slightly too large. The first part of the third term, $-4s\eta''_1/(i\Omega)^3$, is of special interest, as it represents the sole contribution of the term $\eta''$ in equation (3.613) to the value of $\bar{\eta}$ to this order. The term $\eta''$ represents all that remains in the equations of type $\alpha$ of the terms in B neglected in § 3.3 in obtaining the equations of type $\beta$. The initial value of $-4s\eta''_1/(i\Omega)^3$ is only $3{\cdot}46 \times 10^{-6}$ of the first term in $\bar{\eta}$ in the 50 degrees trajectory, and this ratio does not tend to increase as the velocity diminishes. This makes it likely that the equations of type $\beta$ give an accurate solution in all cases when the initial conditions are those of the particular integral.

Returning to the particular integral, we have shown that the third term is only 0·03 (?) of the first term at the vertex of a 50 degrees trajectory where the velocity is as low as 500 f.s. For a trajectory at still higher elevation the minimum velocity is lower; the value of the first term soon becomes too great for the use of approximations which neglect $1-\cos\delta$, while the third term can no longer be neglected in comparison with the first term. The solution therefore fails when the elevation much exceeds 50 degrees as soon as the velocity has fallen much below 500 f.s. The true degree of approximation given by the expansion can only be obtained in a special case. If the terms in $\eta''$ in equation (3.613), and terms of the solution containing $c'$, &c., arising from the terms in $\zeta$ are neglected, it may be shown that the error of the expansion at any stage is less in numerical value than the last term retained.* Hence the numerical estimates of the third term, obtained above, justify the use of the first term only to obtain an approximate value of the drift at all elevations up to 50 degrees and for the initial part of a trajectory at any elevation.

## PART V.—SUMMARY AND CONCLUSION.

### § 5.0. *Summary of Preceding Results.*

In the earlier parts of this paper we have suggested a tentative set of components for the complete force system acting on a shell moving through air (or other medium), in which this complete system may be assumed to depend at any moment only on the position and velocities of the shell. We have submitted these suggestions to the test of experiment, and found that, so far as we have carried the analysis in this paper, the experiments confirm our suggestions, and provide, when the yaw is small, numerical values for two of the force coefficients ($f_M$ with a probable error of 2 per cent. and $f_L$ with a probable error of 10 per cent.) for velocities up to double the velocity of sound. Rough values for a third coefficient $f_H$ are also provided. It appears probable that the other components (except of course the drag) are much less important, and that values of the yaw up to perhaps 10 degrees may be regarded as small in this connection.

It is convenient to summarize here what we do and do not know about the components of the force system on the shells used in this trial. The values of the drag coefficient $f_R$ may be regarded as known for all velocities at zero yaw. The values of $f_M$ and $f_L$ are roughly known for velocities up to $v/a = 2 \cdot 0$, and values of the yaw less than 10 degrees. From wind channel experiments $f_R$, $f_M$ and $f_L$ are all known for all values of the yaw when $v/a$ is small, and these determinations probably apply so long as $v/a < 0 \cdot 7$. The damping effects are only known roughly, but sufficient is known to estimate how long a shell will take effectively to settle down to a steady state of motion.

On the other hand the variation of $f_R$ with yaw is entirely unknown except from wind channel experiments, and so is the variation of $f_M$ and $f_L$ at values of the yaw greater than 10 degrees. The rate of diminution of the axial spin is unknown and so is the size of the swerve effect, though this latter is not likely to be important.

The variation of $f_R$ with yaw could be studied experimentally by a suitable combination of jump card observations, with the use of the solenoid chronograph to determine as exactly as possible the deceleration of the shell at every point. The values of $f_M$ and $f_L$ for larger values of the yaw could be obtained by a detailed analysis of unstable rounds in which large values of the yaw are realized. A start could be made with the data of the present trial, but we cannot undertake this in this paper.

---

* The equation is now of the first order in $\eta$ only, so that the exact solution may be written down in the form of an integral. By successive integration by parts we obtain the expansion (3.632) together with an integral representing the error after $n$ terms.

In Part III., we have arrived at two separate solutions of the equations of motion of a shell treated as a rigid body, which together cover practically all types of motion which are likely to occur in practical shooting. (We ignore here the case of an unstable shell, since it is of no practical use.) A general solution of the equations of motion of type $\alpha$ has been developed, which applies with sufficient accuracy to the most general type of motion of a shell whose angle of yaw $\delta$ and inclination of the tangents of true and plane trajectories do not exceed (say) $0 \cdot 1$ radian. The solution of the equations of type $\beta$ can be applied with sufficient accuracy to the steady (non-oscillatory) motion of a shell at any angle of yaw. In practice the large angles of yaw ($> 0 \cdot 1$ radian) only occur in the neighbourhood of or beyond the vertex of a high angle trajectory, and by this stage the initial angular oscillations of the shell have been completely damped out so that the condition for the applicability of the solution of type $\beta$ is satisfied. Thus the solutions we have obtained, though theoretically inadequate, are probably sufficient to cover all cases likely to occur in practice.

In order to make use of these solutions to determine the complete motion of a shell, information is necessary as to the complete force system acting on the shell. Our information, as we have seen, is fairly complete for angles of yaw up to 10 degrees, and can be applied to calculate the true trajectory of any shell for which the angle of yaw does not exceed this value, if the loss of spin and increase of drag with yaw can be ignored.

Larger angles of yaw (exceeding 10 degrees) occur in general only as a consequence of the low velocity of the shell near the vertex of a high angle trajectory. The force system is then mainly covered by wind channel observations. The information as to the force system obtained by our methods is thus adequate for the calculation of a complete set of twisted trajectories at all elevations, at any rate for a 3-inch shell.

## § 5.1. *Problems for Further Discussion.*

5.11. *Unstable Rounds.*—We have already mentioned that further information about $f_M$ and $f_L$, at yaws greater than 10 degrees, could be obtained by analysis of the unstable rounds. This requires the application of the exact top equation with a variable value of $\mu$ (§ 3.7) to the discussion. No means of introducing damping effects into these equations has yet been devised. It should, however, be possible to obtain fairly reliable information as to the variation of $f_M$ and $f_L$ with yaw between the angles of 10 degrees and 30 degrees by the analysis of the unstable rounds fired in this trial (Table IV.).

5.12. *Initial Conditions.*—By extrapolating the $\delta$-curve and $\phi$-curve backwards to the gun muzzle ($t = 0$) information may be obtained as to the way in which the projectile leaves the gun, which may prove of value. Owing to the effect of the initial oscillations on the ranging of the shell, it is important to determine whether, in general, the initial disturbance takes place at, or nearly at, the same orientation. Secondly, it is important to determine whether the initial oscillations may be regarded as caused by an impulsive couple whose size is independent of the twist of the rifling. If this is so, the amplitude of the initial oscillations of a shell can be cut down indefinitely by sufficiently increasing the spin. If, however, as appears to be the case from a rough survey of the data of the present trial, the initial circumstances are such that the impulsive couple (or its equivalent) increases in proportion to the twist of the rifling, then no increase of spin will reduce the oscillation below a certain definite limit. This conclusion would be technically important, as in the later stages of flight the spin is always in excess of requirements, and so the initial spin should be kept down as much as possible.

5.13. *Wind Effects.*—In calculating the effect of wind on a shell it is usual to assume that the shell at once turns its nose to the relative wind. This is not strictly correct, and the true angular motion in a wind when the velocity is known at every point can be determined by our theory, since the forces acting on the shell depend only on its motion *relative* to the air. Consider, for example, the special case in which a shell suddenly enters a cross-wind region from a region of still air; it starts its relative trajectory with a yaw $\delta$ given by the equation

$$\tan \delta = w/v,$$

where $w$ is the wind velocity and $v$ the velocity of the shell. At the same time $\delta' = 0$ and $\phi' = 0$. The equations of § 3.6 enable the subsequent motion to be properly traced, and the errors in the usual treatment calculated.

## § 5.2. *Effect of Size and Shape of Shell.*

The jump card trials described in this paper were carried out with shells of two different shapes only. The differences between the two shells may be seen from fig. 6 to be considerable, form A having an ogival head of roughly 2 calibres radius, while form B is of 6 calibres radius. For form B the experiments determine the moment coefficient only, for a single position of the centre of gravity, and give no information as to the cross-wind force. As experiments of this type are expensive and laborious to carry out, it is of importance to examine how far these results may be applied to shells of other shapes and sizes.

From the results of § 1.1 it appears that there is no evidence that the size (represented by the radius $r$ of the shell) enters into any of the factors on which the force coefficients depend, so that the coefficients $f_R$, $f_M$, $f_L$ may be considered as entirely independent of size. It is therefore sufficient to make experiments on shells of as small a calibre as is consistent with obtaining accurate measurements of the jump cards.

With regard to the effect of variation of shape we have very little evidence.

If we compare the moment coefficients $f_M$ for shells of 2 and 6 calibres radius of head, as shown in figs. 4 and 5, it is obvious that the difference is much less marked than the difference between the two curves of $f_R$, and that the two curves of $f_M$ are very nearly of the same shape. No great errors would be introduced by assuming that the values of $f_M$ for the two shells were in a constant ratio. Thus it seems reasonable, until the appearance of evidence to the contrary, to consider that the value of the moment coefficient for any shell can be obtained by multiplying the value, obtained in this experiment, by a constant independent of the velocity. It will then be sufficient to determine the value of this constant at a single velocity, which may even be a low velocity attainable in a wind channel. The value of the cross-wind force factor for any shell may be obtained in a similar manner but the results will be more uncertain. For rough purposes it may even be sufficient to assume $f_L$ and $f_M$ independent of the velocity except when dealing with velocities very near the velocity of sound. It thus appears to be possible to treat $f_M$ in the classical way in which $f_R$ was treated, in which it was assumed that the values of $f_R$ for two different shells are in a constant ratio at all velocities. This treatment is inadequate in the case of $f_R$, but on present evidence is far more valid in the case of $f_M$.

By applying the results of the present trial in this way, we may even hope to get reasonably accurate estimates of the drift and stability for any type of shell, on the basis of wind channel experiments only on the particular shape of shell required. The method would be especially valuable in connection with the design of new shapes of shell. It is known that, in general, the longer and more pointed a shell is, the less is its drag coefficient; by a series of wind channel tests on a series of shell shapes it would be possible to determine the greatest length of shell that would be sufficiently stable in a gun of given rifling, or the sharpness of rifling required to make a given shell stable. Useful information was obtained on this point from wind channel experiments before the jump card trial provided certain data for the extrapolation to high velocities. It must be emphasised, however, that this one experiment needs extension and confirmation before the structure sketched above can confidently be reared upon it.

We have now discussed in general terms the applicability of our theory and experiments to the calculation of drift, stability, the effect of wind, and the design of improved forms of shell. Though the details of the calculations on these various points are not given here, enough has been said to show that the results form some advance in the subject of the application of aerodynamics to the flight of shells.

## TABLE V.—General Table of Results.

Column 1. Number of round.
,, 2. Muzzle velocity, f.s., for round, or mean for group.
,, 3. Air density $\rho$, lb./(ft.)$^3$ and temperature ° F.
,, 4. $\phi' = \frac{1}{2}\Omega$, degrees/sec.
Upper entry—Calculated value for round, or mean for group.
Lower entry—Observed value.*
,, 5. Period T, between successive minima of $\delta$, in units of $10^{-3}$ sec.
,, 6. Maxima of yaw, $\alpha_1(t)$ degrees.
,, 7. Mean values of $\Omega T$, radians, for each round, for mean velocity as stated.
,, 8. Minima† of yaw, $\beta_1(t)$ degrees.

* When there is only one entry there was no detectable difference between the observed and calculated values of $\phi$.

† *Note.*—The sign given is the sign of $\beta_1(t)$ at the minimum, see § 4.0. The yaw $\delta$ is always positive.

Gun rifled 1 turn in 40 calibres.

| 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|
| | | | | Type 1. | | | |
| I. 11 | 918 | 0·0792<br>43° | 2108 | 252<br>267<br>— | 2·1<br>2·5<br>2·1? | 19·10<br>905 | 0·0<br>— |
| I. 12 | 918 | 0·0792<br>43° | 2108 | 276<br>262<br>— | 6·1<br>6·9<br>7·5+ | 19·79<br>905 | −0·4<br>−0·6 |
| I. 14 | 920 | 0·0807<br>42° | 2113 | 296<br>289 | 7·2<br>6·4 | 21·60<br>905 | −0·9<br>−1·5? |
| I. 13 | 931 | 0·0807<br>42° | 2139 | 234<br>260<br>— | 1·5<br>3·1<br>5·5+ | 18·44<br>919 | −1·0<br>−2·4? |
| I. 5 | 1565 | 0·0782<br>45° | 3595 | 230<br>— | 12·2<br>8·5 | 28·87<br>1539 | −0·9 |
| I. 6 | 1565 | 0·0782<br>45° | 3595 | 227<br>— | 13·7<br>9·5 | 28·49<br>1540 | 0·0 |

TABLE V. (continued).

| 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|
| Type I. (continued). | | | | | | | |
| I. 7 | 1565 | 0·0782<br>45° | 3595<br>3375 | 354 | 10·5 | 41·71<br>1526 | −2·3 |
| I. 15 | 2130 | 0·0807<br>42° | 4892 | 2T =<br>213 | —<br>?7·5<br>5·5 | 18·19<br>2082 | —<br>−0·4 |
| I. 16 | 2130 | 0·0807<br>42° | 4892 | 2T =<br>203 | —<br>?7·5<br>5·7 | 17·33<br>2085 | —<br>−0·6 |
| I. 1 | 2167 | 0·0786<br>42° | 4977 | 93½<br>92<br>90½ | 4·5<br>4·1<br>3·7 | 15·98<br>2104 | −1·0<br>−0·6<br>−0·6 |
| I. 2 | 2167 | 0·0786<br>42° | 4977 | 112<br>105<br>— | 7·5<br>6·3<br>5·3 | 18·85<br>2117 | 0·0<br>−0·2 ? |
| I. 3 | 2167 | 0·0786<br>42° | 4977 | 107<br>99½<br>— | 4·5<br>3·5<br>3·1 | 17·94<br>2120 | −0·3<br>−0·4 |
| I. 4 | 2167 | 0·0786<br>42° | 4977 | 118<br>121½ | 4·1<br>3·0 | 20·81<br>2113 | −0·4<br>−1·4 ? |
| I. 19 | 2272 | 0·0812<br>40° | 5217 | 98½<br>90½<br>— | 5·0<br>4·3<br>3·0 | 17·20<br>2217 | −0·2<br>−0·3 ? |
| I. 20 | 2346 | 0·0812<br>40° | 5388<br>5250 | 96½<br>94½<br>— | 8·7<br>7·9<br>7·1 | 17·96<br>2288 | +0·4<br>+1·0 |
| I. 21 | 2346 | 0·0812<br>40° | 5388 | 99<br>99<br>— | 5·6<br>4·0<br>3·3 | 18·03<br>2282 | −1·0<br>−0·0 |

TABLE V. (continued).

| 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|
| Type II. Form A. C.G. forward. | | | | | | | |
| II. 8 | 934 | 0·0807<br>42° | 1960<br>2020 | 210<br>224<br>— | 1·6<br>3·2<br>3·8 | 15·30<br>923 | −0·9<br>−0·5 |
| II. 9 | 934 | 0·0807<br>42° | 1960 | 246<br>228<br>— | 4·5<br>4·0<br>4·7 | 16·22<br>922 | −0·3<br>−1·2 |
| II. 10 | 934 | 0·0807<br>42° | 1960 | 254<br>256<br>— | 6·7<br>6·8<br>8·1 | 17·45<br>921 | −0·8<br>−0·5 |
| II. 5 | 1585 | 0·0780<br>46° | 3541 | 147½<br>145½<br>— | 4·9<br>3·7<br>2·4 | 18·11<br>1554 | −0·7<br>−0·7 |
| II. 6 | 1585 | 0·0780<br>46° | 3541 | 144½<br>145½<br>— | 1·2<br>1·2<br>1·0 | 17·92<br>1555 | −0·5 ?<br>−0·5 ? |
| II. 7 | 1585 | 0·0780<br>46° | 3541 | 187<br>170 | 3·1<br>2·2 | 22·07<br>1548 | −0·4<br>−0·3 ? |
| II. 1 | 2024 | 0·0786<br>42° | 4795<br>4530 | 99<br>89½<br>91½ | 3·2<br>2·5<br>2·3 | 14·76<br>1983 | −0·7<br>−0·8<br>−0·7 |
| II. 2 | 2024 | 0·0786<br>42° | 4795<br>4625 | 103<br>95<br>87 | 3·0<br>2·3<br>1·9 | 15·34<br>1982 | 0·0<br>−0·1<br>−0·5 |
| II. 3 | 2024 | 0·0786<br>42° | 4795<br>4435 | 96<br>89<br>92 | 2·9<br>2·6<br>2·3 | 14·29<br>1984 | −0·2<br>−0·2<br>−0·2 |
| II. 4 | 2024 | 0·0786<br>42° | 4795 | 98½<br>81<br>85 | 1·9<br>1·7<br>1·8 | 14·75<br>1985 | 0·0<br>0·0<br>0·0 |

TABLE V. (continued).

| 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|
| Type III. Form A. C.G. back. | | | | | | | |
| III. 8 | 931 | 0·0807<br>42° | 2450 | 247<br>226<br>— | 5·7<br>5·5<br>5·8 | 20·23<br>919 | -0·4<br>-1·5 |
| III. 9 | 931 | 0·0807<br>42° | 2450 | 254<br>216<br>— | 5·0<br>5·5<br>4·3 | 20·10<br>919 | -0·4<br>-1·3 |
| III. 10 | 931 | 0·0807<br>42° | 2450 | 232<br>224<br>— | 5·2<br>5·5<br>7·8 | 19·50<br>920 | -0·3<br>-1·5 |
| III. 5 | 1583 | 0·0780<br>46° | 4166<br>4100 | 217<br>— | 8·7<br>7·8 | 31·06<br>1556 | -3·7 |
| III. 6 | 1583 | 0·0780<br>46° | 4166 | 196<br>— | 9·0<br>8·2 | 28·52<br>1558 | -3·6 |
| III. 7 | 1583 | 0·0780<br>46° | 4166<br>3980 | 237<br>— | 5·5<br>6·3 | 32·93<br>1553 | -2·9 |
| III. 1 | 2025 | 0·0785<br>43° | 5331 | 125<br>97<br>— | 1·5<br>1·2<br>0·8 | 20·66<br>1994 | —<br>— |
| III. 2 | 2025 | 0·0785<br>43° | 5331 | 113<br>97<br>— | 3·7<br>3·4<br>3·4 | 19·54<br>1995 | -0·7<br>-1·5 |
| III. 3 | 2025 | 0·0785<br>43° | 5331 | 109<br>107<br>— | 2·4<br>1·8<br>1·3 | 20·10<br>1994 | -0·2?<br>-0·5 |
| III. 4 | 2025 | 0·0785<br>43° | 5331 | 106<br>102<br>— | 2·9<br>2·1<br>1·8 | 19·36<br>1994 | 0·0<br>-0·8 |

### Table V. (continued).

Gun rifled 1 turn in 30 calibres.

| 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|
| Type I. Form A. Empty shell. | | | | | | | |
| I. 22 | 1119 | 0·0811 | 3429 | 91½ | 2·1 | 10·07 | +0·1 |
| | | | | 86½ | 1·9 | | 0·0 |
| | | | | 82 | 1·9 | | −0·3? |
| | | 38° | | 86 | — | | — |
| | | | | 79½ | 1·9 | | −0·2 |
| | | | | 87 | 2·0 | 1089 | — |
| I. 23 | 1119 | 0·0811 | 3429 | 79 | 3·0 | 10·34 | +0·4 |
| | | | | 91 | 2·4 | | −0·1 |
| | | | | 87 | 2·2 | | −0·4 |
| | | 38° | | 86 | — | | — |
| | | | | 83½ | 2·0 | | −0·4 |
| | | | | 84½ | 1·8 | 1090 | — |
| I. 24 | 1119 | 0·0811 | 3429 | 85 | 2·8 | 10·27 | +0·4 |
| | | | | 90 | 2·4 | | −0·2 |
| | | | | 86 | 2·2 | | −0·4 |
| | | 38° | | 84½ | — | | — |
| | | | | 84½ | 2·1 | | −0·6 |
| | | | | 84 | 2·1 | 1090 | −0·9? |
| I. 25 | 1326 | 0·0811 | 4061 | 74½ | 2·0 | 9·91 | 0·0 |
| | | | | 71 | 1·8 | | 0·0 |
| | | | | 69 | 1·9 | | 0·0 |
| | | 38° | | 68½ | — | | — |
| | | | | 71 | 1·6 | | 0·0 |
| | | | | 70 | 1·4 | 1280 | −0·8? |
| | | | | — | 2·0 | | |
| I. 26 | 1326 | 0·0811 | 4061 | 71 | 3·6 | 10·04 | +0·4 |
| | | | | 74 | 2·8 | | +0·1 |
| | | | | 71 | 2·6 | | 0·0 |
| | | 38° | | 73 | — | | — |
| | | | | 66½ | 2·6 | | −0·1 |
| | | | | 69½ | 2·2 | 1285 | −0·8? |
| I. 27 | 1563 | 0·0805 | 4786 | 60½ | 3·8 | 9·59 | −0·3 |
| | | | | 55½ | 3·7 | | −0·3? |
| | | | | 59 | 3·5 | | −0·2 |
| | | 36° | | 60½ | — | | — |
| | | | | 55½ | 2·9 | | −0·5 |
| | | | | 56½ | 3·3? | | −0·8 |
| | | | | — | 2·5 | 1515 | |

## TABLE V. (continued).

| 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|
| Type I. Form A. Empty shell (continued). | | | | | | | |
| I. 28 | 1563 | 0·0805 | 4786 | 59 | 3·1 | 9·66 | −0·4 |
| | | | | 59 | 2·6 | | −0·4 |
| | | | | 59 | 2·1 | | −0·3 |
| | | 36° | | 59 | — | | — |
| | | | | 57½ | 1·9 | | −0·2 |
| | | | | 54½ | 2·0 | | −0·5 ? |
| | | | | — | 2·1 | 1515 | |
| Type II. Form A. C.G. forward. | | | | | | | |
| II. 17 | 1119 | 0·0811 | 3168 | 92½ | 2·8 | 9·85 | 0·0 |
| | | | | 90 | 2·75 | | 0·0 |
| | | | | 88½ | 2·6 | | −0·6 |
| | | 38° | 3128 | 91 | — | | −0·5 |
| | | | | 91½ | 2·0 | | −0·7 |
| | | | | — | 1·7 | 1092 | |
| II. 18 | 1119 | 0·0811 | 3168 | 88 | 3·3 | 9·88 | +0·1 |
| | | | | 94½ | 2·8 | | 0·0 |
| | | | | 87½ | 2·6 | | −0·6 |
| | | 38° | 3128 | 89 | — | | −0·7 |
| | | | | 91 | 2·3 | | −0·7 |
| | | | | — | 2·2 | 1093 | |
| II. 19 | 1119 | 0·0811 | 3168 | 88 | 2·9 | 9·82 | +0·1 |
| | | | | 89 | 2·4 | | 0·0 |
| | | | | 89½ | 2·0 | | −0·3 |
| | | 38° | 3128 | 90 | — | | −0·4 |
| | | | | 89½ | 2·0 | | −0·5 ? |
| | | | | 91½ | 2·1 | 1088 | −0·5 ? |
| II. 24 | 1292 | 0·0805 | 3709 | 76 | 2·7 | 9·70 | −0·15 |
| | | | | 74 | 2·9 | | −0·3 |
| | | | | 72 | 2·7 | | −0·2 |
| | | 36° | | 76 | — | | −0·3 |
| | | | | 75 | 2·4 | | −0·7 |
| | | | | 76 | 3·0 | 1259 | −0·5 |
| II. 22 | 1589 | 0·0819 | 4738 | 52 | 2·1 | 8·93 | +0·5 |
| | | | | 51 | 1·9 | | +0·1 ? |
| | | | | 54½ | 1·4 | | — |
| | | 36° | | 56½ | 1·4 | | — |
| | | | | 54½ | — | | +0·5 |
| | | | | 52 | 1·1 | | — |
| | | | | 55½ | 1·4 | 1543 | +0·1 |

## Table V. (continued).

| 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|
| Type II. Form A. C.G. forward (continued). | | | | | | | |
| II. 23 | 1589 | 0·0819 | 4738 | 54 | 2·3 | 9·13 | +0·3 |
| | | | | 5 periods in 276 | 2·1 | | +0·1 |
| | | | | | — | | — |
| | | 36° | | | — | | — |
| | | | | | — | | — |
| | | | | | 1·9 | | — |
| | | | | | 1·6 | 1548 | −0·1 |
| Type III. Form A. C.G. back. | | | | | | | |
| III. 17 | 1119 | 0·0811 | 3928 | 67 | 1·8 | 10·08 | 0·0 |
| | | | | 75 | 1·8 | | 0·0 |
| | | | | 71 | 2·2 | | 0·0 |
| | | 38° | | 73 | 2·2 | | 0·0 |
| | | | | 71 | — | | 0·0 |
| | | | | 74 | 1·8 | | 0·0 |
| | | | | 77 | 1·7 | 1091 | 0·0 |
| III. 18 | 1119 | 0·0811 | 3928 | 63 | 1·4 | 9·91 | +0·2 |
| | | | | 71 | 1·4 | | — |
| | | | | 75 | 1·4 ? | | +0·1 |
| | | 38° | | 77 | 1·8 | | — |
| | | | | 72 | — | | +0·2 |
| | | | | 71½ | 1·4 | | +0·1 |
| | | | | 67 | 1·3 | 1091 | −0·1 |
| III. 19 | 1119 | 0·0811 | 3928 | 64 | 1·6 | 10·17 | +0·1 |
| | | | | 72 | 1·4 | | +0·1 |
| | | | | 77 | 1·3 | | +0·1 |
| | | 38° | | 74 | 1·4 | | +0·1 |
| | | | | 74 | — | | +0·1 |
| | | | | 71 | 1·1 | | −0·1 |
| | | | | 77 | 1·1 | 1091 | — |
| III. 20 | 1292 | 0·0805 | 4534 | 62 | 2·7 | 9·63 | 0·0 |
| | | | | 59 | 2·7 | | −0·1 |
| | | | | 62 | 2·2 | | −0·5 |
| | | 36° | | 58 | 2·2 | | — |
| | | | | 63 | — | | −1·0 |
| | | | | 63 | 2·0 | 1261 | −0·5 |
| | | | | 60 | 1·6 | | −0·9 |
| | | | | — | 1·6 | | |

## TABLE V. (continued).

| 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|
| Type III. Form A. C.G. back (continued). | | | | | | | |
| III. 21 | 1292 | 0·0805 | 4534 | 60 | 3·6 | 9·50 | 0·0 |
| | | | | 62 | 3·2 | | 0·0 |
| | | | | 59 | 2·7 | | -0·2 ? |
| | | 36° | | 59 | 2·6 | | — |
| | | | | 61 | — | | -0·7 ? |
| | | | | 59 | 2·3 | | -0·5 |
| | | | | 60 | 2·1 | 1262 | -0·6 |
| III. 22 | 1567 | 0·0805 | 5501 | 42 | 3·5 | 9·38 | -0·1 |
| | | | | 50 | 2·9 | | -0·2 |
| | | | | 49 | 2·8 | | -0·4 |
| | | 36° | | 49 | 2·4 | | — |
| | | | | 50 | — | | 0·4 |
| | | | | 49 | 2·6 | | -0·3 |
| | | | | 46 | — | | -0·3 |
| | | | | — | 2·3 | 1525 | |
| III. 23 | 1567 | 0·0805 | 5501 | 51 ? | 1·7 ? | 9·22 | 0·0 |
| | | | | 42 | 1·6 | | 0·0 |
| | | | | 51 | 1·4 | | 0·0 |
| | | 36° | | 47 | 1·0 | | — |
| | | | | 48 ? | — | | — |
| | | | | 51 ? | 1·0 ? | | 0·0 |
| | | | | 49 ? | — | 1526 | 0·0 |
| | | | | | 1·0 | | 0·0 |
| Type IV. Form B. | | | | | | | |
| IV. 21 | 900 | 0·0811 | 2431 | 116 | 1·7 | 10·27 | -0·6 |
| | | | | 120 | 2·8 | | -1·5 |
| | | | | 122 | 2·3 | | — |
| | | 38° | | 116 | — | | -1·0 |
| | | | | 126 | 2·7 + | 884 | -2·0 ? |
| | | | | — | 3·5 + | | |
| IV. 22 | 900 | 0·0811 | 2431 | 124 | 3·0 | 10·10 | +0·3 |
| | | | | 128 | 2·8 | | -0·2 |
| | | | | 119 | 1·9 | | — |
| | | 38° | | 110 | — | | -0·4 |
| | | | | 119 | 2·3 | 884 | -0·7 |
| | | | | — | 3·0 + | | |

## TABLE V. (continued).

| 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|
| Type IV. Form B (continued). | | | | | | | |
| IV. 23 | 900 | 0·0811 | 2431 | — | 1·7 ? | 9·84 | +0·2 |
| | | | | 112 | 1·7 | | −0·8 |
| | | | | 120 | 2·6 | | — |
| | | 38° | | 114 | — | | −0·8 |
| | | | | 118 | 2·0 | 884 | −1·0 |
| | | | | — | 2·0+ | | |
| IV. 13 | 1078 | 0·0811 | 2911 | 113 | 2·0 | 12·39 | +0·3 |
| | | | | 128 | 1·8 | | −0·4 |
| | | | | 125 | 1·5 ? | | −0·6 |
| | | 38° | | 113 | 1·6 | | −0·9 |
| | | | | — | 2·0 | 1059 | |
| IV. 14 | 1078 | 0·0811 | 2911 | 109 | 1·6 | 11·45 | +0·2 |
| | | | | 123 | 1·5 | | −0·3 |
| | | | | 110 | 1·8 ? | | — |
| | | 38° | | 105 | 1·6 | | −0·3 |
| | | | | — | 1·7 | 1060 | |
| IV. 15 | 1078 | 0·0811 | 2911 | 107 | 2·2 | 11·76 | +0·3 |
| | | | | 121 | 2·1 | | −0·1 |
| | | | | 118 | 2·2 | | — |
| | | 38° | | 108 | 1·7 | | −0·5 |
| | | | | — | 1·8 | 1060 | |
| IV. 16 | 1547 | 0·0811 | 4178 | 82 | 1·6 | 10·72 | +0·1 |
| | | | | 78 | 1·6 | | 0·0 |
| | | | | 67 | 1·4 | | — |
| | | 38° | | 75 | 1·0 | | 0·0 |
| | | | | 74 ? | 0·9 | 1503 | −0·1 |
| IV. 17 | 1547 | 0·0811 | 4178 | 83 | 3·5 | 10·72 | +0·1 |
| | | | | 75 | 3·2 | | 0·0 ? |
| | | | | 75 | 2·8 | | — |
| | | 38° | | 73 | 1·9 | | −1·2 ? |
| | | | | 71 | 2·6 | 1503 | −0·9 ? |
| IV. 18 | 1547 | 0·0811 | 4178 | 83 | 4·2 | 11·09 | +0·1 |
| | | | | 80 | 3·9 | | 0·0 |
| | | | | 75 | 3·3 | | — |
| | | 38° | | 73 | 3·2 | | −0·4 |
| | | | | — | 3·1 | 1509 | −1·1 |

## Table V. (continued).

| 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|
| Type IV. Form B (continued). | | | | | | | |
| IV. 19 | 1547 | 0·0811 | 4178 | 77 | 4·0 | 10·96 | |
| | | | | 75 | 3·4 | | — |
| | | 38° | | 73 | 3·4 | 1496 | |
| IV. 20 | 1547 | 0·0811 | 4178 | 73 | 3·3 | 10·69 | |
| | | | | 75 | 2·5 | | — |
| | | 38° | | 72 | 2·2 | 1497 | |
| IV. 24 | 2101 | 0·0805 | 5675 | 55 | 4·6 | 10·70 | −0·2 |
| | | | | 57 | 3·8 | | −0·3 |
| | | | | 51 | 3·5 | | — |
| | | 36° | | 55 | 3·5 | | −0·5? |
| | | | | 53 | 3·4 | 2045 | −1·0? |
| IV. 25 | 2112 | 0·0805 | 5705 | 56 | 2·2 | 10·70 | 0·0? |
| | | | | 55 | 2·2 | | 0·0 |
| | | | | 52 | 1·3 | | — |
| | | 36° | | 55 | 1·2 | 2073 | 0·0? |
| | | | | 54 | 1·3 | | −0·8 |
| IV. 26 | 2149 | 0·0805 | 5805 | 50 | 2·4 | 10·60 | −0·1 |
| | | | | 53 | 2·0 | | 0·0 |
| | | | | 54 | 2·0 | | — |
| | | 36° | | 53 | 1·7 | | — |
| | | | | 54 | 1·2 | 2093 | −0·3 |

## TABLE VI.—Values of the Stability Coefficient and the Air Couple deduced from Analysis of the *Stable* Shells.

Summary of Notation used in the Headings of this Table.

$\mu \sin \delta$ = couple due to air forces.

$s = \dfrac{\Omega^2 B}{4\mu}$ = stability coefficient.

$v$ = mean velocity of shell, f.s.

$\rho$ = air density, lb./(ft.)$^3$.

$a$ = velocity of sound, f.s.

$f_M(v/a) = \mu/(\rho v^2 r^3)$, the air couple coefficient.

N.B.—The values of $s$, $\mu$ and $f_M$ have not been corrected in this table for the effect of the cards.

| 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|
| Round No. or group of rounds. | Twist of rifling. | Value of $s$ deduced from observation. | Mean value of $v$ corresponding to value of $s$. | Value of $\mu$. | Value of $f_M(v/a)$. | Value of $v/a$. | Total percentage spread of $s$ or $\mu$ in group. |
| I. 11, 12 | 1/40 | 1·113 | 905 | 1220 | 9·58 | 0·824 | 4·1 |
| I. 14 | 1/40 | 1·087 | 906 | 1250 | 9·71 | 0·825 | |
| I. 13 | 1/40 | 1·131 | 919 | 1230 | 9·26 | 0·837 | |
| I. 22–24 | 1/30 | 1·61 | 1090 | 2230 | 11·85 | 0·996 | 3·2 |
| I. 25, 26 | 1/30 | 1·66 | 1283 | 3030 | 11·62 | 1·173 | 1·7 |
| I. 27, 28 | 1/30 | 1·74 | 1515 | 4000 | 11·09 | 1·388 | 1·1 |
| I. 5– 7 | 1/40 | 1·005 | 1535 | 3920 | 10·89 | 1·394 | 0·9 |
| I. 15, 16 | 1/40 | 1·137 | 2084 | 6410 | 9·36 | 1·897 | 6·7 |
| I. 1 | 1/40 | 1·180 | 2104 | 6390 | 9·40 | 1·916 | |
| I. 2– 4 | 1/40 | 1·118 | 2117 | 6750 | 9·80 | 1·927 | |
| I. 19 | 1/40 | 1·152 | 2217 | 7190 | 9·22 | 2·023 | 1·9 |
| I. 20, 21 | 1/40 | 1·133 | 2285 | 7800 | 9·41 | 2·085 | |
| II. 8–10 | 1/40 | 1·172 | 922 | 1170 | 8·71 | 0·840 | 4·9 |
| II. 17–19 | 1/30 | 1·69 | 1091 | 2000 | 10·46 | 0·997 | 0·8 |
| II. 24 | 1/30 | 1·72 | 1259 | 2700 | 10·86 | 1·153 | |
| II. 5– 7 | 1/40 | 1·121 | 1553 | 3680 | 9·98 | 1·408 | 4·7 |
| II. 22, 23 | 1/30 | 1·94 | 1546 | 3800 | 9·95 | 1 416 | 3·8 |

## TABLE VI. (continued).

| 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. |
|---|---|---|---|---|---|---|---|
| Round No. or group of rounds. | Twist of rifling. | Value of $s$ deduced from observation. | Mean value of $v$ corresponding to value of $s$. | Value of $\mu$. | Value of $f_M(v/a)$. | Value of $v/a$. | Total percentage spread of $s$ or $\mu$ in group. |
| II. 1 | 1/40 | 1·216 | 1983 | 5506 | 9·12 | 1·804 | |
| II. 2 | 1/40 | 1·200 | 1982 | 5689 | 9·42 | 1·803 | |
| II. 3 | 1/40 | 1·232 | 1984 | 5320 | 8·80 | 1·805 | |
| II. 4 | 1/40 | 1·220 | 1985 | 5809 | 9·59 | 1·806 | |
| III. 8–10 | 1/40 | 1·107 | 919 | 1520 | 11·41 | 0·837 | 0·9 |
| III. 17–19 | 1/30 | 1·64 | 1091 | 2570 | 13·66 | 0·997 | 3·3 |
| III. 20, 21 | 1/30 | 1·76 | 1232 | 3200 | 12·79 | 1·156 | 2·1 |
| III. 22, 23 | 1/30 | 1·84 | 1526 | 4500 | 12·32 | 1·398 | 2·9 |
| III. 5– 7 | 1/40 | 1·035 | 1556 | 4410 | 11·96 | 1·411 | 0·6 |
| III. 1– 4 | 1/40 | 1·109 | 1994 | 7020 | 11·52 | 1·814 | 1·4 |
| IV. 21–23 | 1/30 | 1·64 | 884 | 1270 | 10·25 | 0·808 | 2·6 |
| IV. 13–15 | 1/30 | 1·39 | 1060 | 2140 | 12·06 | 0·969 | 5·0 |
| IV. 16–20 | 1/30 | 1·505 | 1502 | 4070 | 11·41 | 1·373 | 3·7 |
| IV. 24 | 1/30 | 1·525 | 2045 | 7420 | 11·29 | 1·874 | 1·0 |
| IV. 25 | 1/30 | 1·525 | 2075 | 7500 | 11·11 | 1·899 | |
| IV. 26 | 1/30 | 1·54 | 2093 | 7680 | 11·16 | 1·917 | |

TABLE VII.—Observed Values of $h+\kappa$ and $h-\kappa+2\gamma$, the Damping Factors for each Group.

Groups fired at a velocity of 900 f.s. apparently have negative damping and are not included.

| Group.<br>Muzzle velocity, f.s. | Calculated value of $\kappa$. | $h+\kappa$. | $h-\kappa+2\gamma$. | Probable value of $h$. |
|---|---|---|---|---|
| I. 22–24<br>1119 | 0·4 | 1·9 | 1·6 ? | 1·8 |
| I. 25, 26<br>1326 | 0·3 | 2·4 | 1·2 | 1·8 |
| I. 27, 28<br>1563 | 0·4 | 3·0 | 0·5 | 1·8 |
| I. 1– 4<br>2167 | 0·7 | 2·2 | 0·6 | 1·4 |
| I. 19–21<br>2320 | 0·8 | 2·2 | −0·2 ? | 1·3 |
| II. 17–19<br>1119 | 0·4 | 2·2 | 1·2 | 1·7 |
| II. 24<br>1292 | 0·2 | 0·9 | 0·6 | 1·5 |
| II. 5– 7<br>1585 | 0·4 | 3·4 | 0·4 | 2·0 |
| II. 22, 23<br>1589 | 0·4 | 3·3 | 0·6 | |
| II. 1– 4<br>2027 | 0·6 | 3·0 | 0·6 | 2·0 |
| III. 17–19<br>1119 | 0·4 | 0·7 ? | 0·1 ? | 1·0 |
| III. 20, 21<br>1292 | 0·2 | 3·1 | 0·9 | 2·0 |
| III. 22, 23<br>1567 | 0·4 | 3·0 | 0·3 | 2·0 |
| III. 1– 4<br>2025 | 0·6 | 4·2 | 1·7 | 3·0 |
| IV. 13–15<br>1078 | 0·5 | 0·7 | 1·4 | 1·0 |
| IV. 16–18<br>1547 | 0·5 | 3·1 | 1·2 | 2·0 |
| IV. 24–26<br>2120 | 0·7 | 5·0 | 0·9 | 3·0 |

N.B.—The calculated value of $\kappa$ is obtained by using the value of the cross-wind force coefficient given in Table I.

### TABLE VIII.—Plane Trajectories at 50 degrees and 30 degrees, with Calculations of the Drift, &c., for Shells of External Form A.

Constants used in the Calculations.

| | | |
|---|---|---|
| Muzzle velocity | | 2000 feet per second. |
| Centre of gravity | | 4·88 inches from the base |
| Weight | | 16·02 lb. |
| Moments of inertia | A | 0·1329 lb. (ft.)$^2$. |
| | B | 1·1555 lb. (ft.)$^2$. |
| $\Omega$ for | gun rifled 1 in 30 | 192·7 radians per second. |
| | „ 1 „ 40 | 144·5 „ „ |
| Ballistic coefficient. | | 1·75. |

### TABLE VIIIA.—Plane Trajectories at Elevations of 50 degrees and 30 degrees.

Column 1. Time $t$ seconds.
„ 2. Velocity $v_1$, feet per second.
„ 3. Inclination $\theta_1$, degrees.
„ 4. Horizontal distance X, feet.
„ 5. Vertical height Y, feet.

| | Elevation 50 degrees. | | | | Elevation 30 degrees | | | |
|---|---|---|---|---|---|---|---|---|
| 1. | 2. | 3. | 4. | 5. | 2. | 3. | 4. | 5. |
| | | ° ′ | | | | ° ′ | | |
| 0 | 2000 | 50 0 | 0 | 0 | 2000 | 30 0 | 0 | 0 |
| 1 | 1720 | 49 21 | 1,199 | 1,413 | 1726 | 29 8 | 1,614 | 917 |
| 2 | 1506 | 48 36 | 2,254 | 2,628 | 1515 | 28 8 | 3,033 | 1692 |
| 3 | 1342 | 47 44 | 3,201 | 3,686 | 1352 | 26 59 | 4,300 | 2354 |
| 4 | 1218 | 46 45 | 4,068 | 4,624 | 1230 | 25 42 | 5,454 | 2926 |
| 6 | 1059 | 44 28 | 5,648 | 6,241 | 1075 | 22 45 | 7,538 | 3867 |
| 8 | 959 | 41 48 | 7,116 | 7,619 | 985 | 19 24 | 9,455 | 4608 |
| 10 | 877 | 38 43 | 8,514 | 8,805 | 916 | 15 42 | 11,264 | 5182 |
| 12 | 807 | 35 13 | 9,857 | 9,819 | 860 | 11 40 | 12,988 | 5603 |
| 14 | 746 | 31 15 | 11,155 | 10,671 | 816 | 7 19 | 14,639 | 5881 |
| 16 | 693 | 26 45 | 12,411 | 11,369 | 780 | 2 43 | 16,227 | 6021 |
| 18 | 647 | 21 44 | 13,631 | 11,921 | 751 | − 2 7 | 17,755 | 6030 |
| 20 | 609 | 16 10 | 14,817 | 12,330 | 728 | − 7 5 | 19,228 | 5912 |
| 24 | 557 | 3 40 | 17,098 | 12,738 | 703 | − 17 11 | 22,017 | 5314 |
| 28 | 537 | − 9 50 | 19,266 | 12,624 | 700 | − 26 55 | 24,608 | 4261 |
| 32 | 547 | − 22 55 | 21,331 | 12,013 | 712 | − 35 51 | 27,011 | 2790 |
| 36 | 579 | − 34 25 | 23,293 | 10,930 | 735 | − 43 40 | 29,228 | 938 |
| 40 | 627 | − 43 54 | 25,151 | 9,403 | | | | |
| 44 | 681 | − 51 30 | 26,902 | 7,466 | | | | |
| 48 | 735 | − 57 32 | 28,539 | 5,157 | | | | |
| 52 | 786 | − 62 23 | 30,057 | 2,521 | | | | |
| 55 | 818 | − 65 25 | 31,114 | 359 | | | | |

TABLE VIIIB.—Calculation of the Drift, Stability, and Damping Factors, for the Gun Rifled 1 in 30.

Column 1. The time $t$, seconds.
,, 2. The stability factor $s$.
,, 3. $-4s\theta'_1/\Omega$.
,, 4. $4s\kappa/\Omega$.
,, 5. $\psi$, equation (4.203).
,, 6. The drift Z, feet.
,, 7. The azimuth, arc tan (Z/X), degrees.
,, 8. The second term in the expansion of $\bar{\eta}$, equation (3.632), given by

$$\frac{\eta^{(2)}}{(i\Omega)^2} = \frac{4s}{(i\Omega)^2}\left\{\frac{d}{dt}(4s\theta'_1) + 4s\kappa\theta'_1 + c'\int_0^t 4s\kappa\theta'_1\,dt/c\right\}.$$

,, 9. The damping factor (§ 4.22),

$$(\sigma_0/\sigma)^{\frac{1}{2}}\,e^{-(q_1-q_0)}.$$

Elevation 50 degrees.

| 1. | 2. | 3. | 4. | 5. | 6. | 7. | 8. | 9. |
|---|---|---|---|---|---|---|---|---|
| 0 | 1·945 | 0·00042 | 0·0217 | 0 | 0 | | | 1 |
| 1 | 2·598 | 0·00062 | | 0·00038 | 0·2 | | | 0·702 |
| 2 | 3·334 | 0·00098 | 0·0214 | 0·00082 | 0·8 | | | 0·520 |
| 3 | 4·189 | 0·00139 | | 0·00129 | 1·8 | | | 0·404 |
| 4 | 5·129 | 0·00192 | 0·0221 | 0·00182 | 3·1 | | 0·00007 | 0·330 |
| 6 | 7·123 | 0·00321 | | 0·00405 | 7·5 | 0 5 | | 0·196 |
| 8 | 10·95 | 0·00569 | 0·0570 | 0·00752 | 15·9 | | 0·00066 | 0·115 |
| 10 | 14·60 | 0·00868 | | 0·0114 | 29·1 | 0 12 | | 0·076 |
| 12 | 18·39 | 0·01240 | 0·0518 | 0·0155 | 47·3 | 0 16 | 0·00155 | 0·055 |
| 14 | 22·20 | 0·01693 | | 0·0199 | 70·4 | | | 0·043 |
| 16 | 26·33 | 0·0227 | 0·0585 | 0·0250 | 98·6 | 0 27 | 0·00334 | 0·034 |
| 18 | 30·88 | 0·0297 | | 0·0308 | 133 | | | |
| 20 | 35·16 | 0·0372 | 0·0663 | 0·0374 | 173 | | 0·0056 | |
| 24 | 42·63 | 0·0511 | | 0·0530 | 276 | 0 55 | | |
| 28 | 45·87 | 0·0562 | 0·0755 | 0·0705 | 409 | 1 13 | | |
| 32 | 43·24 | 0·0486 | | 0·0885 | 573 | 1 32 | | |
| 36 | 37·11 | 0·0353 | 0·0700 | 0·1050 | 762 | | 0·0012 | |
| 40 | 30·11 | 0·0231 | | 0·1192 | 971 | 2 12 | | |
| 44 | 24·02 | 0·0147 | 0·0595 | 0·1315 | 1190 | | | |
| 48 | | | | 0·1420 | 1415 | | −0·0019 | |
| 52 | | | | 0·1510 | 1637 | | | |

## TABLE VIIIB. (continued).

Elevation 30 degrees.

| 1. | 2. | 3. | 4. | 5. | 6. | 7. |
|---|---|---|---|---|---|---|
| | | | | | | ° ' |
| 0 | 1·945 | 0·00056 | | 0 | 0 | |
| 1 | 2·55 | 0·00086 | | | | |
| 2 | 3·21 | 0·00122 | | 0·00080 | 1·1 | |
| 4 | 4·79 | 0·00234 | | 0·00174 | 4·1 | |
| 6 | 8·43 | 0·00368 | | 0·00351 | 9·5 | |
| 8 | 9·11 | 0·00581 | | 0·00655 | 19·2 | 0 7 |
| 10 | 11·16 | 0·00785 | | 0·01022 | 34·4 | |
| 12 | 14·01 | 0·01063 | | 0·01417 | 55·3 | 0 15 |
| 14 | 15·75 | 0·01276 | | 0·01828 | 82·0 | |
| 16 | 17·48 | 0·01495 | | 0·0225 | 114·3 | 0 24 |
| 18 | 18·85 | 0·01671 | | 0·0269 | 152·0 | |
| 20 | 19·92 | 0·01810 | | 0·0316 | 195·0 | 0 35 |
| 24 | | 0·0191 | | 0·0418 | 297 | 0 46 |
| 28 | | 0·0174 | | 0·0525 | 419 | 0 59 |
| 32 | | 0·0143 | | 0·0629 | 557 | 1 11 |
| 36 | | 0·0110 | | 0·0729 | 708 | 1 23 |

## TABLE IX.—Comparison of Calculated Drift with Observations of April–May and February, 1918.

The azimuth of the shell at time $t$ (in minutes of angle) = $At$.

Elevation 50 degrees.

Observations of April–May.

| Rifling 1/30. | | | Rifling 1/40. | | |
|---|---|---|---|---|---|
| Mean observed time. | Mean observed A. | Calculated A. | Mean observed time. | Mean observed A. | Calculated A. |
| 10·9 | 1·46 | 1·27 | 10·2 | 1·18 | 0·90 |
| 23·9 | 2·24 | 2·29 | 22·9 | 1·30 | 1·66 |
| 33·3 | 2·89 | 2·85 | 31·0 | 1·95 | 2·10 |
| 41·3 | 3·16 | 3·36 | 39·1 | 2·00 | 2·44 |

## TABLE IX. (continued).

### Elevation 50 degrees (continued).

### Observations of February.

| Rifling 1/30. | | | Rifling 1/40. | | |
|---|---|---|---|---|---|
| Mean observed time. | Mean observed A. | Calculated A. | Mean observed time. | Mean observed A. | Calculated A. |
| 6·99 | 4·54 | 0·86 | 6·33 | 3·61 | 0·63 |
| 15·03 | 1·68 | 1·65 | 14·07 | 1·40 | 1·18 |
| 26·08 | 1·99 | 2·46 | 24·93 | 1·47 | 1·86 |

### Elevation 30 degrees.

### Observations of April–May.

| Rifling 1/30. | | | Rifling 1/40. | | |
|---|---|---|---|---|---|
| Mean observed time. | Mean observed A. | Calculated A. | Mean observed time. | Mean observed A. | Calculated A. |
| 10·04 | 0·975 | 1·05 | 9·58 | 1·63 | 0·79 |
| 20·6 | 1·575 | 1·77 | 19·35 | 0·80 | 1·28 |
| 27·9 | 2·09 | 2·06 | 25·95 | 1·04 | 1·50 |

### Observations of February.

| Rifling 1/30. | | | Rifling 1/40. | | |
|---|---|---|---|---|---|
| Mean observed time. | Mean observed A. | Calculated A. | Mean observed time. | Mean observed A. | Calculated A. |
| 13·2 | 1·40 | 1·32 | 13·02 | 1·73 | 0·97 |
| 22·52 | 1·36 | 1·86 | 22·05 | 1·25 | 1·38 |

# XI. *A Selective Hot-Wire Microphone.*

*By* W. S. Tucker, *D.Sc.*, *A.R.C.Sc.*, and E. T. Paris, *M.Sc.* (*Lond.*).

*Communicated by Prof.* H. L. Callendar, *F.R.S.*

Received November 2, 1920,—Read January 20, 1921.

CONTENTS.

| § | | Page |
|---|---|---|
| 1. | Introduction | 389 |
| 2. | Description of the Microphone | 391 |
| 3. | Electrical Connections | 394 |
| 4. | Sharpness of Tuning of the Microphone | 396 |
| 5. | Sensitivity | 400 |
| 6. | The Resistance Changes in the Wire Grid | 404 |
| 7. | Experiments on the Measurement of Sound | 411 |
| 8. | The Effect of Tilting the Microphone | 416 |
| 9. | Some Observations of Distribution and Intensity of Sound made with the Hot-Wire Microphone | 423 |
| 10. | Summary | 429 |

## § 1. *Introduction.*

The instrument described in the following paper provides :—

(i) A convenient means of detecting a note of given pitch when other sounds are present ; and

(ii) A method of estimating the relative intensities of sounds of the same pitch.

The idea which formed the starting-point for the construction of the instrument—viz., the placing of an electrically heated grid of fine platinum wire in the orifice of an otherwise closed vessel—was originally employed by one of us (W. S. T.) in the construction of a sound-detector for the use of Sound Ranging Sections in the British Army.* In its original form, the detector was intended to respond to heavily damped aerial vibrations, such as those produced by the firing of guns. Further experiments, however, showed that the detector could be tuned to respond to any continuous sound of definite frequency by suitably choosing the dimensions of the vessel and its orifice.

The tuned instrument is highly selective in its action. It is very sensitive when used to detect low-pitched sounds, but its sensitivity is diminished for the higher pitches. The highest note with which we have experimented hitherto was one of 512 vibrations per second.

---

* British Patent No. 13123 of 1916, and No. 8948 of 1918 ; United States Patent No. 269902 of 1919.

During the course of a large number of experiments with various types of sound-collectors and transmitters, we have found this selective form of hot-wire microphone to be of great assistance. It is very simply constructed and easily manipulated, and for many purposes the only electrical circuit needed is a Wheatstone's Bridge. If, however, it is desired to use aural methods, or if the sound to be observed is exceedingly faint, it is necessary to amplify (by means of thermionic valves) the electrical effects occurring in the microphone. When amplification is used it is possible to detect and render audible a pure tone which is quite inaudible to the unaided ear.

This form of microphone provides us with a very convenient instrument for comparing the efficiencies of various forms of sound-collector—particularly when these are considered in relation to the wave-length of the sound employed. It can also be used for determining the distribution of intensity at the focus of an acoustical lens or mirror, and (what is very important in many practical problems) the manner in which sound is diffracted by obstacles of various shapes and sizes. In addition to these and similar experiments, the microphone can be employed to estimate the relative strengths of the harmonics in an impure sound such as that produced by the usual form of Seebeck's siren. Some examples of these applications of the microphone will be given in the last section of this paper.

So far as we are aware, there is no other instrument of a selective character which could be used for making observations of the kind indicated above. In nearly all cases where attempts have been made to measure or analyse sounds, the instruments employed have depended on the setting in vibration of some form of diaphragm. Such instruments are generally insensitive to notes of moderately low pitch, and are, moreover, easily disturbed by vibrations communicated through the mounting of the diaphragm. For this reason methods of amplification are often of little service if the mountings are to be moved during the experiment.

The hot-wire instrument here described seeks to avoid this disadvantage by measuring directly vibrations which are set up in the air itself, but the displacements in progressive waves are so extremely small that they have been increased by resonance. This employment of resonance naturally limits the scope of the microphone (so that it cannot, for example, be employed for telephony), but it has the advantage not only of magnifying the sound to be recorded, but also of isolating from a complex sound the particular tone which it is desired to measure.

The closed vessel with a single orifice (in which the platinum wire grid is mounted) forms the well-known Helmholtz resonator. The advantages possessed by this form of resonator are :—

(1) That the resonance is sharp.
(2) That the overtones are all relatively high.

(3) That the overtones are not in harmonic relation ; and

(4) That the dimensions of the resonator need only be small compared with the wave-length of the sound to be observed.

The simplified theory of such resonators is due to **Rayleigh**,[*] who showed that the number of vibrations in the resonant note is given by

$$N = \frac{V}{2\pi}\sqrt{\frac{c}{S}},$$

where V is the velocity of sound in the gas in the neck, S is the volume of the reservoir, and $c$ is a quantity depending on the shape and dimensions of the orifice and called (from an electrical analogy) the " conductivity " of the orifice.

### § 2. *Description of the Microphone.*

The complete microphone, comprising the Helmholtz resonator with the platinum wire grid suitably mounted in the neck, is made for convenience in three separate parts :—

(i) The platinum wire grid, mounted in a circular mica plate.

(ii) The " holder," which includes the neck of the resonator and the necessary contact-pieces and terminals for carrying current to the grid ; and

(iii) The " container," or reservoir.

A short description of each of these three parts will now be given.

(i) *The Platinum Wire Grid.*—Fig. 1, A, shows one form of the grid. It consists of a circular plate of thin mica 4 cms. in diameter, in the centre of which is cut a circular hole 0·65 cm. in diameter. A number of small pin-holes are punched at the edge of

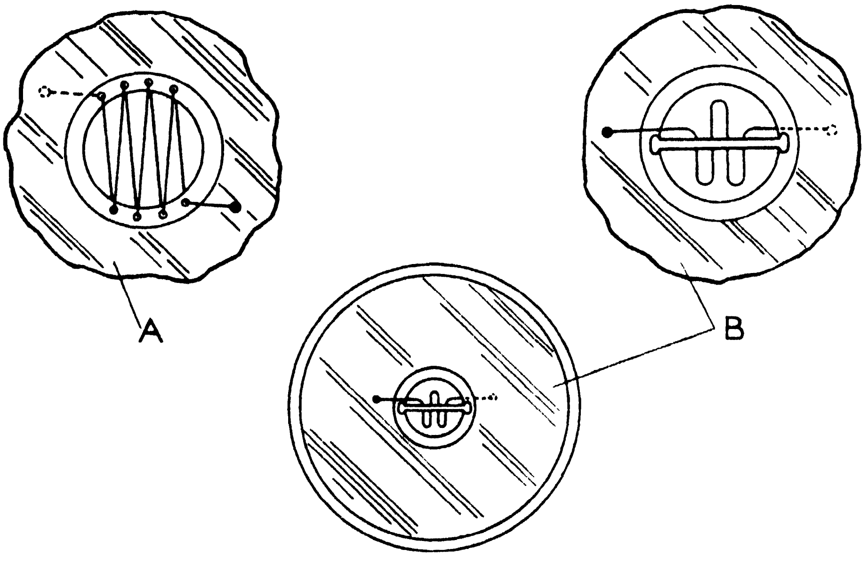


Fig. 1.

---

[*] **Rayleigh**, 'Theory of Sound,' vol. II., p. 174.

this circular opening, and through these the wire is threaded to form a zig-zag grid as shown in the figure. To each side of the mica plate is attached an annular plate of silver foil (shown by shaded portion in the figure), and to each of these plates is soldered one end of the platinum wire. The two plates of silver foil thus constitute the electrodes of the wire grid.

Fig. 1, B, shows another form of grid, for the design and manufacture of which we are indebted to the Research Department of the General Post Office. The wire is bent into three loops, and is supported by a small rod of glass-enamel placed diametrically across the opening in the mica plate.

The grids are made in the first place with Wollaston wire, the silver sheath being removed by means of nitric acid after the wire has been mounted in position. During this part of the process of manufacture the silver foil electrodes are protected by a coating of paraffin wax.

The grids used in the experiments described in this paper were, unless otherwise stated, of the type shown in fig. 1, B. The wire, the diameter of which was about 0·0006 cm., carried a maximum current of about 30 milliamperes, the exact amount varying in individual grids according to the sample of wire used in their manufacture. The average resistance at 10° C. was about 140 ohms, and about 350 ohms when carrying a safe working current of 25 to 28 milliamperes. In the case of a particular grid, it was found that its resistance was 125·5 ohms at 0° C., 133 ohms when carrying a current of 0·5 milliampere, 156·2 ohms at 100° C., and 332 ohms when carrying its working current of 32 milliamperes. The working current heats the grid to just below red heat.

For certain purposes grids were made from wire 0·0015 cm. in diameter and carrying a maximum current of about 55 milliamperes. The number of loops was increased in these grids to eight, four on each side of the glass-enamel support.

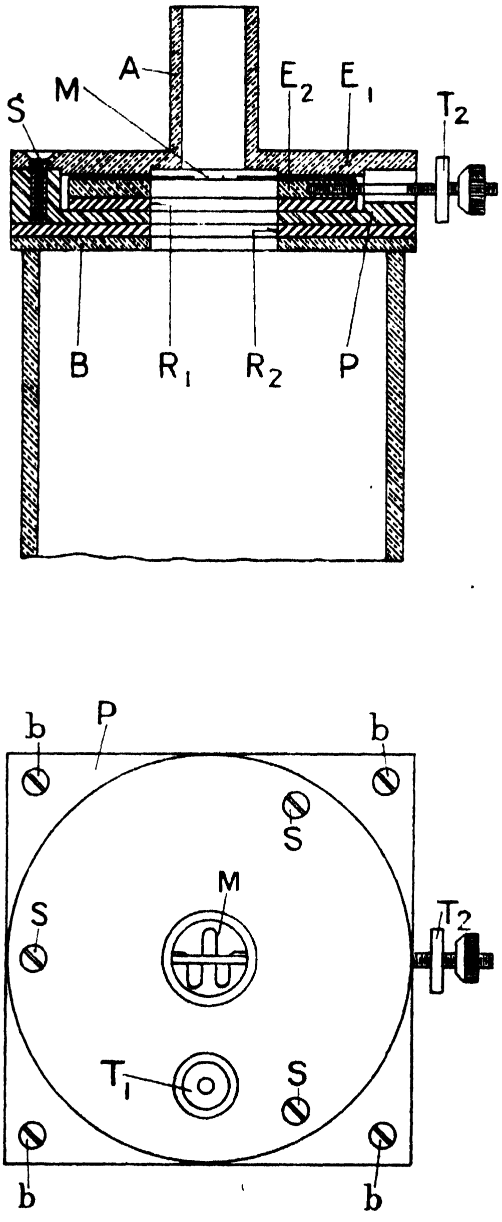


Fig. 2.

(ii) *The "Holder."*—The manner in which the holder is made up is shown in fig. 2. The cylindrical neck A, made of brass, is soldered into the centre of the circular plate

$E_1$ made of the same material. $E_1$ is provided with the terminal $T_1$. The mica plate (M) carrying the grid is clamped between $E_1$ and the lower ring $E_2$, which is also of brass and carries the terminal $T_2$ at the side of the holder. Beneath $E_2$ is a rubber ring $R_1$. and this rests on a bed of ebonite (P) to which also the plate $E_1$ is fixed by the screws S. The ebonite bed (P) is square, and is bolted at the corners to the square brass plate (B) which forms one end of the container. To ensure an air-tight joint a square plate of thin rubber $R_2$ is inserted between the holder and the container.

When the plate $E_1$ is screwed down on to the ebonite bed, so that the mica plate with its silver foil electrodes is firmly held between $E_1$ and $E_2$, a current can be passed through the grid by connecting a battery to the terminals $T_1$ and $T_2$.

The neck (A) forms the channel of communication between the interior of the container and the outside air, and from an acoustical point of view is the most important part of the holder. If the capacity of the container be given, it is on the hydrodynamical conductivity of this neck that the pitch of the resonator depends. The dimensions of the neck generally used were : length 2·2 cms., internal diameter 0·75 cm. In certain experiments, however, the neck was made rather shorter than this in order to tune the resonator to some given pitch.

When the grid is being placed in position between $E_1$ and $E_2$ it is important to see that the circular aperture in the mica plate is coaxial with the neck, since even a small displacement from this position will change the pitch of the resonator by an appreciable amount.

(iii) *The "Container."*—The containers were in most cases made from brass tubing. One end of the tubing is closed with a circular brass plate, while at the other end is fitted a square brass plate of the same dimensions as the base of the ebonite bed of the holder, which is bolted to it by means of the bolts *b*, as shown in fig. 2. A circular hole ½-inch in diameter is cut in the middle of this square plate to allow a free passage of air through the neck into the container. The thickness of the brass of which the tubing was made was 1 mm.

The natural pitch of the resonator of course depends on the volume of the container. Thus, with the form of neck described above, a volume of 290 c.c. gives the resonator a pitch of 116 vibrations per second, while a volume of 68 c.c. gives it a pitch of 240 vibrations per second. For pitches below 200 vibrations per second it has been found convenient to use brass tubing from 2 to 2½ inches in diameter, while for higher pitches (above 200) tubing about 1 inch in diameter is the most suitable.

Other forms of container have been made and tested, and reference to some of these will be found in a later paragraph. The material from which the container is made, and the thickness and rigidity of the walls, have a very marked effect on its resonating properties. The most efficient resonator which was tested was one which had been drilled out of a solid piece of brass, and its superiority must be attributed, in the main, to the increased strength of the walls.

For experimental purposes it is often desirable to have a microphone whose pitch can be varied at will. This can easily be made by fitting a wooden plunger inside the brass tubing of the container, or by making the container in two parts of different sized tubing so that one part will slide over the other.

Another method of tuning the microphone is to alter the length of the neck. It is, however, inadvisable to reduce the length of the neck to less than 1 cm., as with shorter necks than this the grid is exposed to the effect of transient currents of air.

## § 3. *Electrical Connections.*

When the air in the neck of the resonator is set in vibration by a sound of suitable frequency, the platinum wire grid suffers a change in resistance, which may be regarded as being made up of an oscillatory change and a steady change. There are thus two ways in which the microphone can be used.

(i) *The Amplifier Method.*—If the oscillatory effect is to be observed it is necessary to include an amplifier in the circuit. A suitable form for the circuit to take is shown in fig. 3,

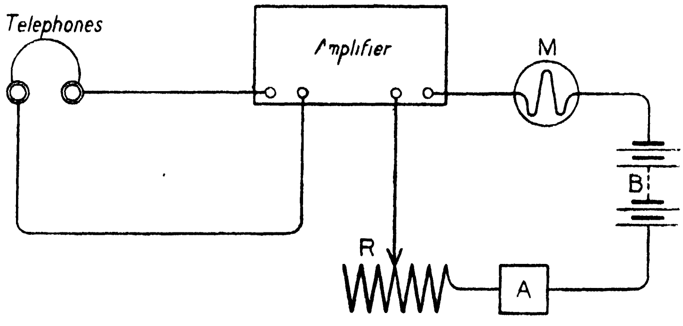


Fig. 3.

where the microphone (M) is connected in series with a battery (B), a milliammeter (A), a rheostat (R), and the primary of the input transformer of a three-valve amplifier.* The sound can of course be heard in the telephones, and provided that the grid lies in an approximately horizontal plane (*i.e.*, that the axis of the neck of the resonator is vertical) the pitch heard is the same as that of the original sound. The effects which occur when the grid is moved out of the horizontal plane are described in § 8 ; but it may be noted here that not only the pitch of the note heard in the telephones, but also the sensitivity of the microphone, depend on the inclination of the grid to the horizontal plane. It is therefore important in any experiment where comparisons of the strength of sound are being attempted, that the position of the microphone relatively to a horizontal plane should not be changed during the experiment. The difficulty can be overcome by arranging that the microphone shall always hang so that the axis of the neck is vertical. Small deviations of one or two degrees from this position do not materially affect the sensitivity.

---

* Several different types of amplifier have been tried for this purpose. The best results have been obtained with a four-valve resistance amplifier specially designed for low acoustic frequencies.

The grid carries, normally, a heating current of about 27 milliamperes. When the resonator responds to a sound, the to-and-fro motion of the air in the neck produces, as already stated, an oscillatory change of resistance, and the effect of this is to superimpose on the steady heating current a ripple of small amplitude (generally only a few microamperes). It is this ripple which is amplified and which is heard in the telephones.

It is shown in a later paragraph (§ 7) that the magnitude of the amplified current may be used as a means of estimating the amplitude of a sound. For this purpose the telephones are replaced by a vibration galvanometer tuned to the pitch of the sound.

This method of employing the microphone for the measurement of a sound, however, is not altogether satisfactory, on account of the difficulty of maintaining an amplifier in such a condition that the current amplification is constant for any length of time; and for this reason the Wheatstone's Bridge method is sometimes preferable. The advantages of the Amplifier Method are that it is very sensitive (especially when a vibration galvanometer is used) and that the microphone can be placed in a moving piece of apparatus, subject only to the restriction that its axis must always be vertical (or at some fixed angle to the horizontal). Vibrations communicated through the mounting of the microphone (even when they are produced by striking the container) have very little effect on the sound heard in the telephones.

(ii) *The Wheatstone's Bridge Method.*—This method is preferable to the Amplifier Method on account of its greater simplicity, and because there is no danger of the sensitivity changing during the course of a long series of observations. A convenient form for the Bridge to take is shown diagrammatically in fig. 4. The microphone (M) with the milliammeter (A) forms one arm of the Bridge. The balancing arm (R) is made about equal to the resistance of the grid when carrying its working current, *i.e.*, about 350 ohms. The rheostat (Rh) is inserted (as shown) in series with the battery, a balance being obtained by adjusting (Rh) until the current through the microphone brings its resistance (together with that of the milliammeter) up to R. For some purposes it is convenient to have a small variable resistance $\rho$ in series with the microphone. In most experiments it is sufficient to take the deflection of the galvanometer as a measure of the intensity of the sound affecting the microphone; but other methods can of course

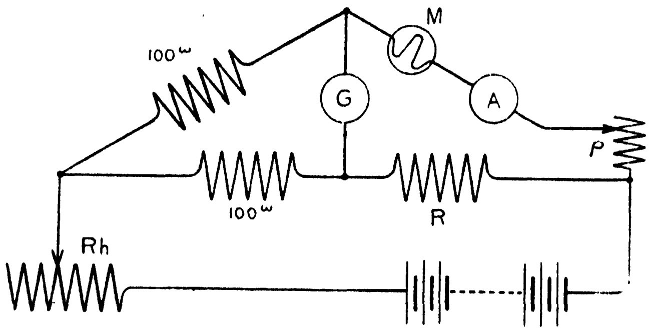


Fig. 4.

be used, such as measuring the increase of current required to bring the grid back to its initial resistance, or determining the alteration in resistance when the current is maintained at a constant value.

It is important when using the microphone in this way that it should not be moved during the course of an experiment. This is one of the disadvantages of the method. A small alteration in the tilt of the microphone upsets the balance of the Bridge and renders the sound-measurements inaccurate. If $\theta$ is the angle between the axis of the microphone and a vertical line (so that $\theta = 0$ when the microphone is held in its normal position with neck uppermost), then it is found that as $\theta$ is increased the resistance of the grid gradually falls and reaches a minimum when $\theta$ is about 100 degrees. The fall in resistance is then about 3 ohms (see fig. 13).

The resistance of the grid also changes when the microphone is rotated about its own axis, except of course when the axis is vertical. Thus, if the microphone is held in a horizontal position and in such a way that the glass-enamel support also lies horizontally, then a rotation of 90 degrees, bringing the glass-enamel support into a vertical position, is accompanied by a fall in the resistance of the grid of about 1 ohm.

All these effects are due to the influence of the convection currents issuing from the heated wire, and it appears that if the convection current from one part of the wire impinges on another part of the wire the resistance of the grid as a whole is always lowered.* As will be seen from the experiments described later, these convection currents play a very important part in the working of the microphone.

Although the resistance of the grid is changed (current being constant) when the plane in which it lies is altered, there is but little change in the sensitivity of the microphone, whether it is held horizontally or vertically, provided that its initial resistance is the same.

## § 4. *Sharpness of Tuning of the Microphone.*

The natural pitch of a microphone can best be determined by plotting its resonance curve. For this purpose the microphone to be tested is set up at a distance of two or three feet from a siren (a modified form of Seebeck's siren was used in the present experiments), and the grid is connected into one arm of a Wheatstone's Bridge as shown in fig. 4. The strength of the blast of air in the siren having been adjusted to a suitable value, a series of readings are taken of the deflection of the galvanometer and the pitch of the siren note. The curve formed by plotting deflection against the interval $n/p$ ($n/2\pi$ being the resonant pitch of the microphone and $p/2\pi$ the pitch of the siren note) gives what we shall call the " resonance curve " of the microphone.

---

* The mutual action due to convection of two electrically heated fine platinum wires is described by J. S. G. THOMAS: " An Electrical Hot-Wire Inclinometer," 'Proc. Phys. Soc. Lond.', vol. XXXII, pp. 291–314 (1920).

A typical example of a curve obtained in this way is shown in fig. 5, the natural pitch of the microphone being 240 vibrations per second.

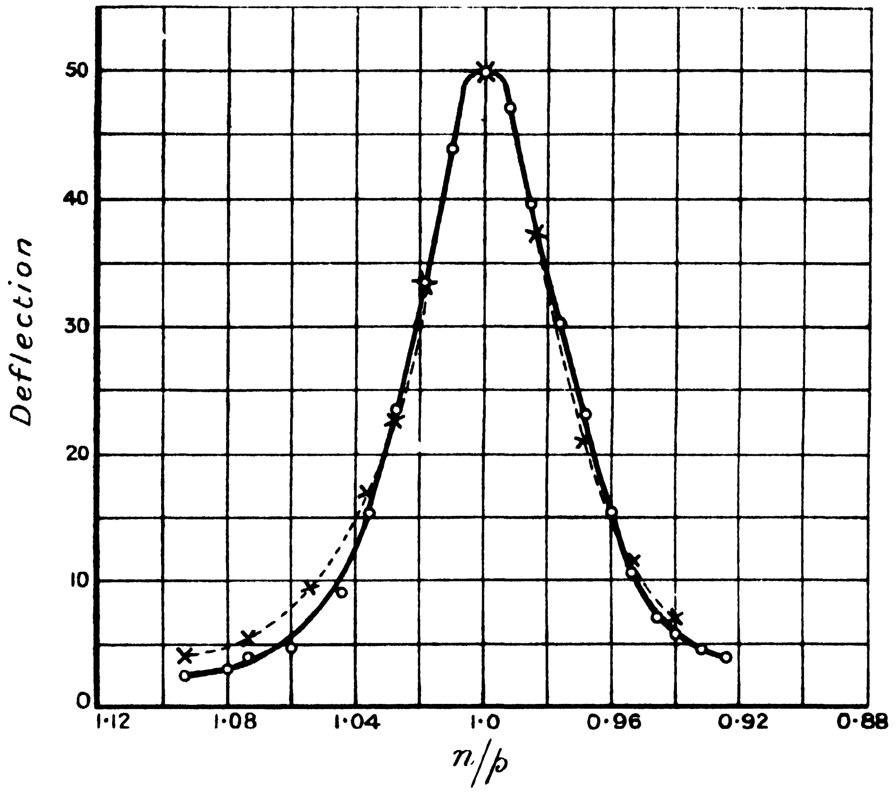


Fig. 5.

In order to obtain reliable resonance curves it was found necessary to make observations out of doors. When the experiments were performed indoors the results were in nearly all cases vitiated by the setting up of stationary waves in the room containing the apparatus.

Experiments are described in a later paragraph (§ 7) which show that—within limits—the change of resistance of the grid is proportional to the square of the amplitude (and therefore to the energy) of the vibration in the neck of the resonator, when the pitch of the stimulating sound remains constant. The influence of a change in the pitch of the sound upon the resistance change of the grid (apart from its effect on the response of the resonator) has not yet been investigated. In dealing with the resonance curves, where in any particular experiment we are concerned only with a comparatively narrow range of frequencies, we shall regard the deflection of the galvanometer as being proportional to the vibrational energy in the neck. Precautions were of course taken to ensure that the deflections were proportional to the changes in resistance suffered by the grid.

As stated above, the source of sound used in these experiments was a modified form of Seebeck's siren. It consisted of a heavy circular brass plate pierced with a ring of twelve equidistant holes. This plate was rotated by an electric motor, the speed of which could be regulated by means of a rheostat in series with the armature. A stream of air, forced through a nozzle against the ring of holes, was supplied from a gas-compressor. The speed of the siren-plate was given by an Elliott speed-indicator attached to the motor, the readings of this instrument being proportional to the frequency of the fundamental note of the siren. The end of the nozzle and the holes in the siren-plate were so shaped that the area through which air could escape from the nozzle was proportional to $(1-\cos pt)$, $p/2\pi$ being the number of holes passing the end of the nozzle per second.* It was found by experiment (using a Hot-Wire Microphone) that with this arrangement the note produced by the siren was remarkably free from harmonics. It will be assumed in what follows that, within the limited range of frequencies over which a resonance curve is plotted, the amplitude of the sound produced by the siren remained sensibly constant.

On the understanding that the results may be subject to some revision on account of these assumptions, we can deduce the degree of damping of the Helmholtz resonator used to obtain the curve in fig. 5.

If the equation of motion of the forced vibration in the neck of the resonator is written

$$\frac{d^2x}{dt^2}+2\mathrm{K}\frac{dx}{dt}+n^2x = f\cos pt,$$

where $\frac{dx}{dt}$ is the instantaneous current of air in the neck, and $n^2 = \mathrm{V}^2c/\mathrm{S}$, then the forced vibration is

$$x = \frac{f}{\{(n^2-p^2)^2+4\mathrm{K}^2p^2\}^{\frac{1}{2}}}\cos(pt-\theta),$$

and the average energy of the vibration in the neck is proportional to the average value of $\left(\frac{dx}{dt}\right)^2$, that is, to

$$\frac{f^2}{n^2\left(\frac{n}{p}-\frac{p}{n}\right)^2+4\mathrm{K}^2}.$$

It is found that the experimental curves can be fairly well represented by an expression of this type, provided that a suitable value is given to K. Thus, by choosing K = 38·5, we obtain the dotted curve shown in fig. 5, which approximates closely enough to the experimental curve to show that this is about the proper value for the damping factor. In all cases so far examined the value of K required to fit the experimental curves has been found to lie between 20 and 40, and it has generally been found that K is less for the low-pitched resonators. For example, with a resonator whose natural pitch was 112 vibrations per second, the value of K was found to be 22·2.

---

* For the design of this siren-plate we are indebted to Messrs. R. H. Fowler and E. A. Milne, of Trinity College, Cambridge.

The forced vibrations of a resonator, due to an external source of sound, have been considered by Rayleigh ("Theory of Sound," vol. II., p. 195). If the periodic change of pressure at the mouth of the resonator is represented by $Fe^{ipt}$, the equation of motion applicable to the forced vibration in the neck is

$$\frac{d^2x}{dt^2}+\frac{p^2c}{2\pi V}\frac{dx}{dt}+n^2x=\frac{cF}{\rho}e^{ipt},$$

where $c$ is the hydrodynamical conductivity of the neck and $\rho$ is the density of the air. The term representing dissipation is here a function of frequency, but it represents "only the escape of energy from the vessel and its neighbourhood, and its diffusion in the surrounding medium, and not the transformation of ordinary energy into heat." It is found to be quite inadequate to account for the experimentally determined rate of dissipation. If $p/2\pi = 112$ vibrations per second, $c = 0\cdot13$ cm. (determined experimentally), and $V = 33760$ cm. per second, then

$$\frac{p^2c}{4\pi V}=0\cdot15,$$

which must be compared with the experimentally determined value of 22·2. It is clear, therefore, in the case of resonators such as those used in these experiments, that the dissipation is due in the main to other causes than the escape of energy through the neck, such as the effect of viscosity on the motion in the neck, and the lack of rigidity in the walls of the container. When we consider the obstructions caused by the glass-enamel rod supporting the grid and the sharp edge of mica at the base of the neck, the comparatively high rate of dissipation is not altogether surprising.

The expression for the natural frequency of a Helmholtz resonator (calculated without allowance for dissipation) is

$$N=\frac{V}{2\pi}\sqrt{\frac{c}{S}}.$$

If N is found from the resonance curve and S is measured, the conductivity $c$ can be calculated, and this should be a constant for a given size and shape of neck. For the cylindrical necks 2·2 cms. long and 0·75 cm. in diameter, and partially obstructed by the platinum wire grid, it is found that $c$ is about 0·13 cm. The following is an example of the kind of measurement taken :—

| N. vibrations/sec. | S. c.c. | c cm. |
|---|---|---|
| 240 | 68 | 0·133 |
| 235 | 73·6 | 0·138 |
| 140 | 197 | 0·131 |
| 116 | 290 | 0·132 |

Temperature 13°·3 C. Mean value of $c = 0\cdot133$ cm.

As a rough check on these observations, we may calculate from hydrodynamical principles approximate upper and lower limits to the conductivity of the neck. The required expression is given by RAYLEIGH ("Theory of Sound," vol. II., p. 181). For a cylindrical neck of length L and radius R,

$$c = \frac{\pi R^2}{L+\alpha R},$$

where $\alpha R$ is the "end correction" to be added to L on account of both ends. Since one end is flanged and the other unflanged, we take $\alpha = 0{\cdot}8 + 0{\cdot}6 = 1{\cdot}4$. To find an upper limit to $c$, take $L = 2{\cdot}2$ cms., $2R = 0{\cdot}75$ cm.; and for the lower limit take $L = 2{\cdot}2$ cms., $2R = 0{\cdot}65$ cm. (the diameter of the circular hole in the mica plate). We then find

$$0{\cdot}162 > c > 0{\cdot}125.$$

## § 5. *Sensitivity.*

The following experiment gives some idea of the smallness of the sound of which the microphone is capable of recording when it is used in conjunction with an amplifier. A microphone was constructed from brass-tubing 1 inch in diameter and tuned to respond to a note of 256 vibrations per second. This microphone was placed in one corner of a large field and connected by a long pair of leads to an army amplifier of the "C Mark II." pattern, the output terminals of which could be connected either to a pair of Brown telephones or to a Campbell vibration-galvanometer (also tuned to 256 vibrations per second). In order to test the sensitivity of the microphone, a tuning-fork giving a note of frequency 256 was sounded over a resonator at various distances from the microphone. The sound produced was as a rule inaudible to the unaided ear at distances greater than 80 yards. It could, however, be heard in the telephones up to distances of about 200 yards, and when the vibration-galvanometer was used quite well-marked deflections were obtained up to a distance of 400 yards or more.

As regards the conditions which determine the sensitivity of the microphone, these may be divided into two groups according to whether they have reference to the resonator or to the grid. We will deal first with those that refer to the resonator. Amongst them the most important is obviously that the resonator be accurately tuned to the note which it is required to record. The effect of mistuning is clearly shown by the resonance curve in fig. 5.

When the resonator is accurately tuned, *i.e.*, when $n = p$, the maximum velocity in the neck, according to the equations given in § 4, is

$$\left(\frac{dx}{dt}\right)_{\text{max.}} = \frac{f}{2K}.$$

Since $f = \frac{cF}{\rho}$, F being the variable part of the pressure at the mouth of the resonator, the actual "magnification" of the displacements and velocities obtained by using the resonator is seen to be

$$\frac{Vc}{2K},$$

and is determined by the ratio of the conductivity of the neck to the damping factor K. Since, when $c$ is constant, K is found to be larger the higher the pitch of the resonator employed, the efficiency of the microphone for the higher notes is correspondingly diminished.

As a numerical example, we may quote the case of the resonator used to obtain the curve in fig. 5, for which $N = 240$ and $K = 38·5$, so that, putting $V = 33760$ cm./sec. and $c = 0·13$ cm., the "magnification" is about 57; while for a resonator of pitch 112, and $K = 22·2$, the "magnification" is about 100. It may be noted here for future reference that if $K = 38·5$, and the amplitude of the sound outside the resonator is $1·27 \times 10^{-7}$ cms., *i.e.*, RAYLEIGH'S value for the minimum amplitude audible when $N = 256$, then the maximum velocity in the neck of the resonator will be about 0·0116 cm. per sec., and that, even if the amplitude is two hundred times the above value, the maximum velocity in the neck will still be less than 2·5 cms. per sec.

One of the most important factors in determining the efficiency of a resonator is the rigidity of the walls of the container. This was well shown by the following experiment.

A cylindrical resonator of rolled veneer was tested and found to respond to the same frequency as a brass resonator of the same volume with the same orifice. The resonant note was 79 vibrations per second. Experiment showed that its degree of response (measured with a Wheatstone's Bridge) was only one-third of that of the brass one, the conditions being as nearly as possible the same in both cases. The resonance curve for the veneer resonator showed that the appropriate value of K was about 35.

We have next to consider some points in connection with the sensitivity of the grid. Almost the first problem that arises in constructing a microphone of this pattern is the choice of a suitable diameter for the wire. In the first experiments that were made with microphones of this type the diameter of the wire used was 0·0015 cm. It was found, however, that better results were obtained with finer wire, and from time to time experiments have been carried out with wire of various diameters down to 0·0002 cm. These experiments showed that the finer the wire the greater was the sensitivity (more especially for high-pitched notes), but that the increased sensitivity obtained with very fine wires was very often counter-balanced by their extreme fragility, which rendered them unsuitable for anything but very special purposes. Finally, a wire of diameter about 0·0006 cm. has been adopted as being sufficiently sensitive, and at the same time not too fragile to prevent its being employed in ordinary out-of-door experiments.

The sensitivity is most easily controlled by altering the heating current. No matter in what manner the microphone is employed, it is found that its sensitivity is always increased by increasing the working current. The curves in fig. 6 show this effect in

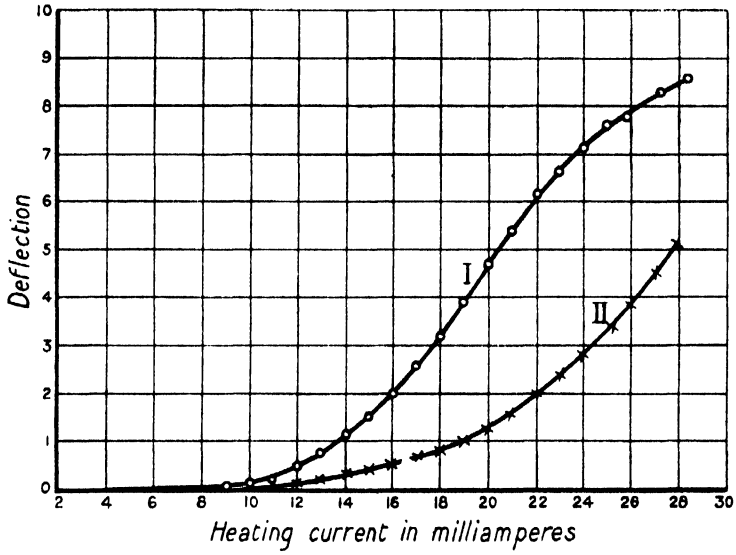


Fig. 6.

two cases : (1) when the microphone is connected in series with the primary of a transformer, the secondary of which is joined to a vibration-galvanometer ; and (2) when the microphone is employed with an amplifier and vibration-galvanometer.

The curves were obtained by clamping the microphone so that its orifice lay just between the prongs of an electrically maintained tuning-fork making 250 vibrations per second. The tuning-fork was carefully maintained at a constant amplitude while the heating current of the microphone was gradually increased from zero to the maximum safe current of about 28 ·5 milliamperes. In fig. 6 the heating current is plotted against the deflection of the vibration-galvanometer. Curve I refers to the case when the microphone is used with a transformer alone, while the effect of introducing an amplifier is shown by Curve II.* In the latter case, however, it must be borne in mind that the amplification itself is in all probability a function of the magnitude of the effect produced in the microphone. It will be observed that the effect produced by the sound is almost negligible until the heating current reaches a value of nearly one-third of the safe maximum.

In the case of the Wheatstone's Bridge, the effect of a change in the heating current of the microphone is complicated by the altered sensitivity of the Bridge. The variation of the sensitivity was therefore investigated by measuring the change in resistance of the grid with various heating currents for a given constant value of sound intensity. The method adopted in the experiment was as follows. A microphone was connected into a Wheatstone's Bridge circuit in the usual manner, and its resistance measured with various heating currents so that a current-resistance curve could be plotted. An electrically maintained tuning-fork, with a resonator to reinforce the sound, was then set in vibration at a convenient distance from the microphone. The amplitude of vibration of the tuning-fork could be observed through a microscope, and it was found that, with care, its amplitude could be maintained at a given value to within 2 or 3 per cent. A second series of observations of current and resistance of microphone grid was then made and the new current-resistance curve plotted on the same chart. The results obtained in a particular experiment are given in the following table:—

---

* For convenience the sensitivity of the galvanometer was reduced when using the amplifier.

| Without Sound. | | With Sound. | |
|---|---|---|---|
| Current in milliamperes. | Resistance in ohms. | Current in milliamperes. | Resistance in ohms. |
| 10·0 | 157·5 | 11·0 | 157·5 |
| 13·6 | 177·5 | 14·9 | 177·5 |
| 16·1 | 197·5 | 17·9 | 197·5 |
| 18·5 | 217·5 | 20·3 | 217·5 |
| 20·3 | 237·5 | 22·6 | 237·5 |
| 22·1 | 257·5 | 24·5 | 257·5 |
| 23·8 | 277·5 | 26·3 | 277·5 |
| 25·2 | 297·5 | 28·0 | 297·5 |
| 26·7 | 317·5 | 29·5 | 317·5 |
| 28·0 | 337·5 | | |

The current-resistance curves are plotted in fig. 7. If, for some particular value of the heating current, the ordinate of the lower curve is subtracted from that of the upper curve, we find the change $\delta R$ which the sound produces in the resistance of the grid

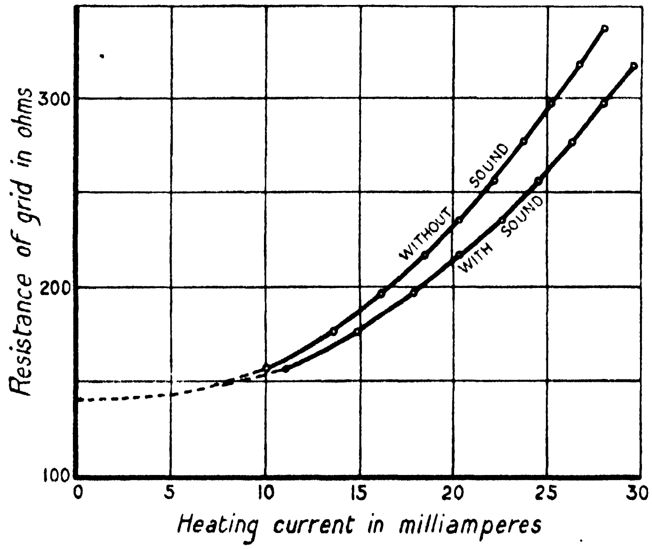


Fig. 7.

under the condition that the current remains constant. That is, $\delta R$ is the change in resistance which would be measured if the bridge were re-balanced by inserting resistance in the microphone arm; or if a bridge of the "constant current" type is used, $\delta R$ is simply proportional to the galvanometer deflection.

The relation between $\delta R$ and R (the initial resistance of the grid), is a linear one, viz.:

$$\delta R = 0{\cdot}2\,(R-140),$$

140 ohms being the resistance of the grid at air temperature. Therefore, by altering R—or, what is the same thing, by altering the heating current—the sensitivity of the grid can be varied in a perfectly definite manner. Conversely, observations which have been made with the same microphone with different heating currents can be very easily made to correspond by reducing them to some standard value of the current. If observations are taken under different conditions of air temperature, a correction on this account can easily be made if desired.

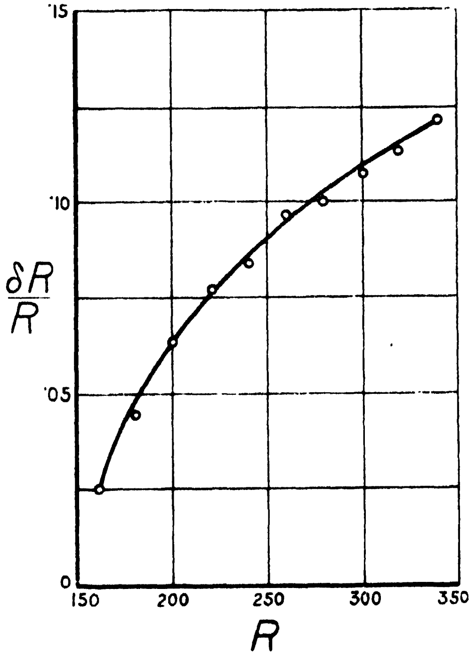


Fig. 8.

A quantity which is more characteristic of the wire from which the grid is made is $\frac{\delta R}{R}$, the change in resistance for a given sound in ohms per ohm. Probably the value of $\frac{\delta R}{R}$ for a given heating current and sound would provide the most convenient method of defining the sensitivity of a grid. The values of $\frac{\delta R}{R}$, obtained from the above table, are plotted against R in fig. 8, and show very clearly the way in which the sensitivity of the grid increases as its temperature rises.

### § 6. *The Resistance Changes in the Wire Grid.*

In this and the following two sections we shall examine more closely the means by which the platinum wire grid is enabled to record electrically the aerial vibrations which are set up in the neck of the resonator. Suppose in the first instance that a microphone is held with its axis vertical (neck uppermost), and that the grid is connected in series with a battery and the primary of the first stage transformer of an amplifier. It is found by experiment that the temperature of the platinum wire, when carrying its normal safe working current of about 29 milliamperes, is in the neighbourhood of 600° C., and we know that in these circumstances the energy supplied to the wire in the form of heat is lost mainly by convection. There is in fact above the grid a free convection current whose velocity depends on the temperature and diameter of the platinum wire. A sound of suitable pitch produces in the neck of the resonator an alternating current of air which is superimposed upon the free convection current, with the result that the convection of heat from the platinum wire is alternately retarded and accelerated. It can easily be seen that if the maximum velocity of the alternating air-current does not exceed the velocity of the convection current, the periodic temperature change produced in the platinum wire will have the same frequency as that of the sound stimulating the resonator.

This is in accordance with the observed fact that when the microphone is held vertically the note heard in the telephones has the same pitch as that of the original sound.

Next, suppose that the microphone is held so that its axis is horizontal and the grid lies in a vertical plane. The free convection current is now at right angles to the axis of the neck, and the effect of an oscillatory motion of the air in the neck (parallel to its axis) will be to produce a periodic change in the temperature (and therefore resistance) of the grid whose periodicity will be twice that of the sound which produces it. This is in fair agreement with observations, for when the microphone is gradually tilted over from a vertical to a horizontal position, the fundamental note heard in the telephone slowly dies away and the octave becomes more and more prominent. The octave is heard best, however, when the neck is pointed slightly downwards so that the axis makes an angle of about 20 degrees with the horizontal. This peculiar effect, which appears to be due to the asymmetrical construction of the neck of the resonator, will be referred to in a later paragraph (§ 8).

We shall in the present section confine our attention to the case when the microphone is held vertically with the neck pointing upwards, and it will be assumed that the resistance changes which the sound produces in the grid can be attributed to the changes in the velocity of the air in the neck. In order to ascertain the nature of the resistance changes which are likely to occur, it is first of all necessary to determine the relation between the resistance R of the grid and the velocity U of the air-current which cools it. U here refers to the velocity of the forced air-current, the actual current passing the wires of the grid being the sum of the forced current and the free convection current. The velocity of the undisturbed convection current rising from the grid will not, of course, be evenly distributed in the neck of the resonator, but the effects of free convection can be represented by an "effective" current $V_0$ supposed uniform throughout the neck. The actual current passing the grid is then $V = U + V_0$, the downward vertical being regarded as the positive direction. We shall first obtain a relation between R and U for small steady values of U, and afterwards extend the results obtained to the case of oscillatory currents by putting $U \sin pt$ in place of U.

On account of the applications to be found in Hot-Wire Anemometry, the cooling of electrically heated fine platinum wires by steady currents of air has received a good deal of attention from physicists. The most complete investigation available is that contained in the well-known paper by L. V. KING.* KING shows that, for fine wires in air-currents of low velocity, the heat-loss per cm. is given by an equation of the form

$$H = \alpha\theta_0/\log\frac{\beta}{Vd},$$

where $\alpha$ and $\beta$ are constants, $\theta_0$ is the excess temperature of the wire above its surroundings, V is the velocity of the air-current, and $d$ the diameter of the wire. This equation is theoretically applicable whenever $Vd < 0 \cdot 0187$, a condition which is amply fulfilled in the present case with $d = 0 \cdot 0006$ cm. and V not greater than 6 cms. per second. We have not, however, been able to adapt this equation in any way which leads to useful results in the case of the Hot-Wire Microphone. It may be remarked that both the diameter of the wire and the magnitude of the air-currents with which we are concerned are considerably smaller than those used in the experiments of KING, or of other investigators to whose work reference will be found in KING's paper. In view of this fact, and of the rather special form of the grid and its mounting, it was thought desirable to determine experimentally the relation between R and U for such small values of U as are likely to be required to account for the behaviour of the grid under the influence of alternating air-currents.

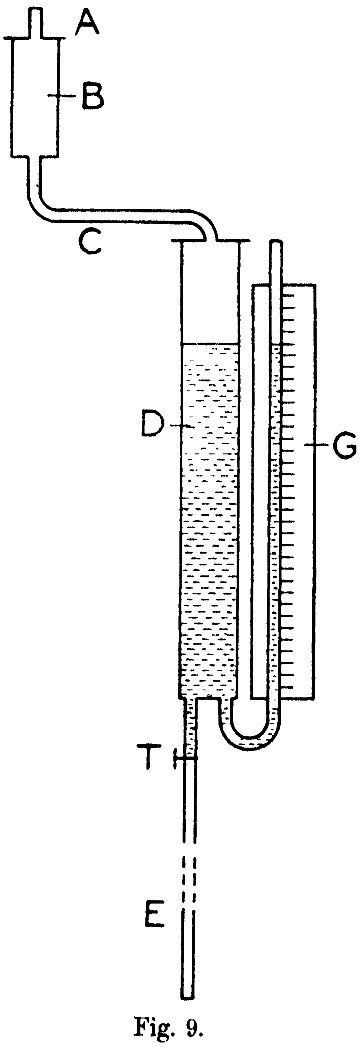


Fig. 9.

The arrangement of the apparatus used in the experiments is shown diagrammatically in fig. 9. A microphone grid is mounted in the holder at A. The interior of the small brass container (B), carrying the microphone and holder communicated by means of the short tube (C) with the reservoir (D), which was partly filled with water. A current of air could be produced past the microphone grid by opening the tap (T), which allowed the water in D to escape through the tube (E) into a second reservoir from 4 to 5 feet below D. A current in the reverse direction could be produced by allowing water to siphon into D from another reservoir at a higher level.

The average velocity of the air-current passing the grid was deduced from a knowledge of the area of the aperture in which the grid lies, the area of the cross-section of D, and the velocity of the fall (or rise) of the surface of the water in D. The latter was found by timing the movement of the level of the water in the gauge (G). The maximum velocity of the fall or rise in D required in the experiments was 0·09 cm. per second, corresponding to a velocity of 6 cms. per second for the air passing the grid.

---

* 'Phil. Trans.,' A, vol. 214, pp. 373–432 (1914).

The microphone formed one arm of the Wheatstone's Bridge shown in fig. 4. The method of taking an observation was as follows. The balancing resistance R was given some prearranged value about equal to the sum of the resistance of the grid (carrying its normal current), the milliammeter and the variable resistance $\rho$ set at 10 ohms. The resistance $\rho$ could be varied by steps of 0·1 ohm from 0 to 20 ohms. With the tap (T) closed, the bridge was then balanced by adjusting the rheostat (Rh), that is, by altering the heating current until the total resistance of the microphone, milliammeter and $\rho$ was equal to R. The tap (T) was then opened and adjusted to give an air-current past the grid of approximately the required velocity. The exact velocity of the air-current was determined by timing with a stop-watch the fall or rise of the level of the water in G for 1, 2 or 3 cm., according to the magnitude of the velocity employed. The change in the resistance of the grid was determined at the same time by increasing or decreasing the resistance $\rho$ so as to restore the balance. The resistance-change was determined to 0·01 ohm by noting the deflection which remained after the bridge had been balanced as nearly as possible by altering $\rho$. The effect of a forced air-current on the resistance of the grid was thus determined under the condition that the electric current carried by the grid remains constant.

The results of one experiment are shown in the following table, which gives U, the impressed air-current, in centimetres per second, and $\delta$R the change in the resistance of the grid. The air-current U is taken as positive when it flows *into* the container—in this case vertically downwards.

MICROPHONE GRID A1079. Heating current 28·5 milliamperes. Resistance of grid when impressed air-current is zero = 270·8 ohms.

| U. cm. per sec. | $\delta$R. ohms. | U. cm. per sec. | $\delta$R. ohms. |
|---|---|---|---|
| −4·12 | −9·73 | +0·27 | +0·46 |
| −3·54 | −7·76 | +0·53 | +0·50 |
| −3·13 | −7·46 | +0·55 | +0·56 |
| −3·04 | −6·83 | +0·90 | +0·73 |
| −2·20 | −4·34 | +1·33 | +0·79 |
| −2·07 | −4·18 | +1·40 | +0·74 |
| −1·65 | −2·91 | +1·97 | +0·47 |
| −1·33 | −2·15 | +1·99 | +0·43 |
| −0·74 | −1·11 | +2·40 | +0·10 |
| −0·48 | −0·76 | +2·67 | −0·37 |
| −0·45 | −0·61 | +3·61 | −1·92 |
| 0 | 0 | +4·30 | −3·36 |

The values of U and $\delta R$ are plotted in fig. 10 and a smooth curve drawn through the points. As the impressed air-current increases from zero the resistance of the grid

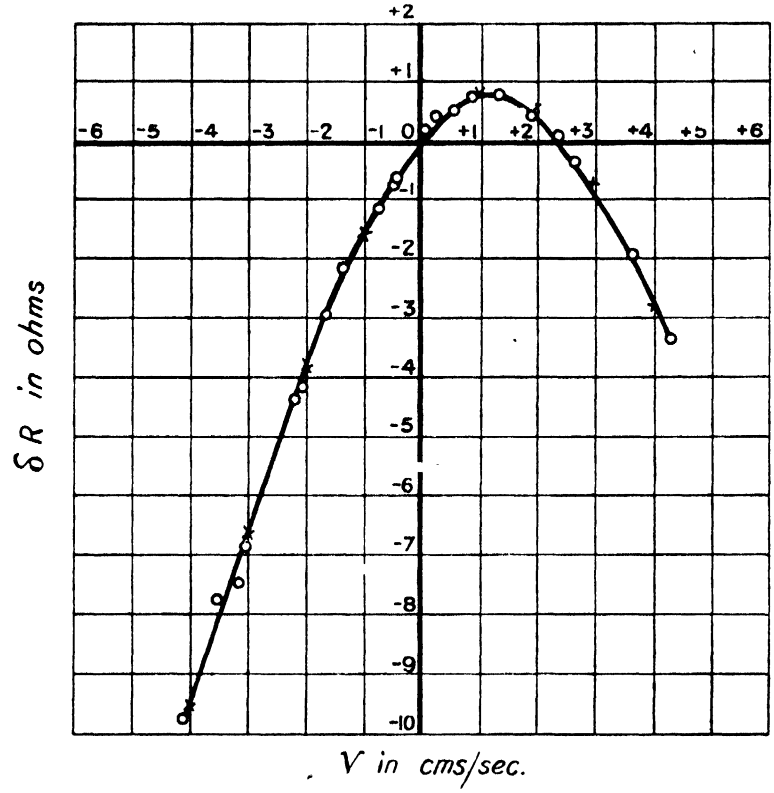


Fig. 10.

(given by $R = 270 \cdot 8 + \delta R$) rises and passes through a maximum, the curve cutting the U-axis at $U = 2 \cdot 45$ cms. per second. The maximum occurs when $U = \frac{1}{2} \times 2 \cdot 45 = 1 \cdot 225$ cms. per second. When U has this value, the impressed air-current balances the free-convection current $V_0$, so that $V_0 = 1 \cdot 225$ cms. per second. Somewhat similar curves to that in fig. 10 are given in a recent paper by J. S. G. THOMAS,* who used this method to determine the velocity of free convection from a platinum wire 0·00784 cm. in diameter and carrying current of 0·6 to 1·2 amperes.

The symmetrical form of the curve about the line $U = V_0 = 1 \cdot 225$ suggests that the relation between $\delta R$ and U can conveniently be represented by a formula of the type

$$\delta R = \delta R_0 + a\,(U - V_0)^2 + b\,(U - V_0)^4 + \&c.,$$

where $\delta R_0$ is the maximum *increase* in resistance occurring when $U = V_0$. It was found that the results of the above experiment could be very fairly represented by the formula

$$\delta R = 0 \cdot 74 - 0 \cdot 50\,(U - 1 \cdot 225)^2 + 0 \cdot 0044\,(U - 1 \cdot 225)^4.$$

A series of points for U = 0, ± 1, ± 2, etc., calculated from this expression, are indicated in fig. 10. In all such experiments, where U did not exceed 5 cms. per second, it was found that the result could be expressed within the errors of the experiment by a formula of the type

---

* 'Phil Mag.', vol. XXXIX, pp. 518-523, and Pl. XI, fig. II (1920).

$$\delta R = \delta R_0 + a(U-V_0)^2 + b(U-V_0)^4.$$

In the above experiment the resistance was measured under the condition that the electric current carried by the grid was the same with and without the air-current. In many experiments, however, the cooling of the grid will be accompanied by an increase of electric current, which tends to restore the temperature to its initial value. The extent to which this takes place depends of course on the particular circuit in which the microphone is used, but in most cases its effect will not be very marked, owing to the dead resistance in series with the microphone.

In a second experiment, performed with a grid of similar type to that used in the first experiment, the resistance was measured for various values of U by rebalancing the bridge with the resistance R, so that the heating current did not remain quite constant but increased or decreased according as the grid was cooled or heated. The resistance Rh was 240 ohms. It was found that the resistance-velocity curve had the same character as that in the first experiment, and up to velocities of 5 cms. per second the change in resistance could be quite adequately represented by an expression of the above type.

It is convenient to write the expression for $\delta R$ in the form

$$\delta R = -2V_0(a+2bV_0^2)U + (a+6bV_0^2)U^2 - 4bV_0U^3 + bU^4.$$

If the values of $a$, $b$ and $V_0$ determined in the above experiment are inserted we get

$$\delta R = 1{\cdot}19U - 0{\cdot}46U^2 - 0{\cdot}022U^3 + 0{\cdot}0044U^4,$$

so that if U is not much greater than 1 cm. per second, $\delta R$ is given by the first two terms to within 2 or 3 per cent.

In applying these results to the case of oscillating air-currents, we shall at first suppose that U is so small that the third and fourth terms are negligible. If it is assumed that the resistance of the grid at any instant is the "equilibrium" value which it would take up if the instantaneous velocity were maintained, then the changes in resistance produced by an alternating air-current U sin $pt$ will be given by

$$\begin{aligned}\delta R &= -2V_0(a+2bV_0^2)U\sin pt \\ &\quad + (a+6bV_0^2)U^2\sin^2 pt \\ &= \tfrac{1}{2}(a+6bV_0^2)U^2 \\ &\quad -2V_0(a+2bV_0^2)U\sin pt \\ &\quad -\tfrac{1}{2}(a+6bV_0^2)U^2\cos 2pt.\end{aligned}$$

The total change would therefore be made up of three parts :---

(1) A steady drop in resistance given by $\delta R_1 = \frac{1}{2}(a+6bV_0^2)U^2$.
(2) A periodic change of resistance $\delta R_2 = -2V_0(a+2bV_0^2)U \sin pt$ of the same frequency as the sound stimulating the resonator.
(3) A periodic resistance change $\delta R_3 = -\frac{1}{2}(a+6bV_0^2)U^2 \cos 2pt$ of frequency twice that of the sound stimulating the resonator.

The relative importance of these effects can be gauged by putting in the values of $a$, $b$, and $V_0$ previously found. This gives

$$\delta R_1 = -0{\cdot}23U^2,$$
$$\delta R_2 = +1{\cdot}19U \sin pt,$$
$$\delta R_3 = +0{\cdot}23U^2 \cos 2pt.$$

These resistance changes correspond to the three most obvious effects of a sound of suitable pitch upon the microphone.

$\delta R_1$ is the effect made use of when the microphone is employed in a Wheatstone Bridge. Since it is proportional to $U^2$, that is, to the energy of the vibration in the neck, it should be proportional to the *intensity* of the sound-wave stimulating the microphone. This is confirmed by the experiments described in § 7.

$\delta R_2$ is the effect which causes the ripple on the heating current, and which can be made audible by the use of an amplifier. It will be seen that the amplitude of this effect is proportional to U, and therefore to the *amplitude* of the sound affecting the microphone. The extent to which this is confirmed by experiment is described in § 7. It should also be noted that the amplitude of the effect is proportional to $V_0$, the free convection current from the grid. It should therefore be possible to increase the loudness of the sound heard in the telephones by artificially increasing the steady air-current passing the grid. That this conclusion is correct can easily be demonstrated by gently heating the brass container with a flame so that a current of air is forced out past the grid.

The existence of $\delta R_3$, which should produce a note in the telephones an octave above the fundamental, is not easy to demonstrate when the microphone is held vertically. It cannot of course be heard in the telephones because it is completely swamped by the fundamental. When, however, the microphone is tilted the octave becomes relatively more important, and is easily heard at certain angles. These effects are described in § 9.

In order to discover what will occur in the case of very loud sounds, it will be necessary to use the more complete expression for $\delta R$ which involves the third and fourth powers of U. Thus, writing the relation between $\delta R$ and U for steady currents in the form

$$\delta R = a'U + b'U^2 + c'U^3 + d'U^4,$$

and substituting U sin $pt$ for U, we find that the total resistance change can now be regarded as being made up of five parts, namely :—

$$
\begin{aligned}
\delta R_1 &= \tfrac{1}{2}b'U^2\left(1+\tfrac{3}{4}\frac{d'}{b'}U^2\right),\\
\delta R_2 &= a'U\left(1+\tfrac{3}{4}\frac{c'}{a'}U^2\right)\sin pt,\\
\delta R_3 &= -\tfrac{1}{2}b'U^2\left(1+\frac{d'}{b'}U^2\right)\cos 2pt,\\
\delta R_4 &= -\tfrac{1}{4}c'U^3\sin 3pt,\\
\delta R_5 &= \tfrac{1}{8}d'U^4\cos 4pt.
\end{aligned}
$$

The interpretation of the various terms is obvious, and it remains only to estimate their relative importance. To do this, we can take as an example the grid examined in the experiment described above and use the numerical values of $a'$, $b'$, $c'$ and $d'$ already given. It will also be supposed that $U = 2\cdot5$ cms. per second, which, as previously shown (§ 5), would be the maximum velocity produced in the neck of the resonator if its natural frequency were 240 vibrations per second, and the amplitude in the primary wave were 200 times as great as the minimum amplitude audible. We find then that

$$
\begin{aligned}
\delta R_1 &= -0\cdot23U^2(1-0\cdot0072U^2) = -1\cdot37,\\
\delta R_2 &= +1\cdot19U(1-0\cdot014U^2) = +1\cdot09\sin pt,\\
\delta R_3 &= +0\cdot23U^2(1-0\cdot0096U^2)\cos 2pt = +1\cdot35\cos 2pt,\\
\delta R_4 &= +0\cdot0055U^3\sin 3pt = +0\cdot086\sin 3pt,\\
\delta R_5 &= +0\cdot00055U^4\cos 4pt = +0\cdot021\cos 4pt.
\end{aligned}
$$

So that, even with a comparatively loud sound, the notes of pitch three and four times the fundamental are quite unimportant.

One other point remains to be noted. From the expressions just given it can be seen that the simple rule, that $\delta R_1$ is proportional to the intensity of the sound stimulating the resonator, does not hold for very loud sounds. Similarly, the amplitude of the fundamental oscillatory effect is not proportional to the amplitude in the primary wave when very intense sounds are used. In both cases the effect with very loud sounds falls short of what it would be if the simple relations continued to hold.

## § 7. *Experiments on the Measurement of Sound*

Two experiments will now be described which were undertaken with the object of testing the correctness of two of the conclusions arrived at in the previous section, viz. :—

(i) That the steady resistance change $\delta R_1$ is proportional to the intensity of the sound affecting the microphone; and

(ii) That the amplitude of the oscillatory resistance change $\delta R_2$ is proportional to the amplitude of the sound which produces it.

As pointed out at the end of § 6, neither of these conclusions would be expected to hold quite exactly in the case of very loud sounds.

(i) *First Experiment.*—The object of the experiment was to find out if the change of resistance $\delta R_1$ is proportional to the intensity of the sound. In order to do this it is only necessary to expose a microphone to different sounds of known relative intensities, and to observe in each case the value of $\delta R_1$. The method adopted in the experiment was to observe the effect produced on the microphone when it was placed at various distances from a source of sound working at a constant rate, the relative intensities of the sound to which the microphone was exposed being deduced from the Inverse Square law.

The source of sound was an electrically maintained tuning-fork vibrating in front of a glass-bottle resonator, the frequency of the fork being 250 vibrations per second. The fork with its resonator was placed on the ground in a suitable open space.* The amplitude of vibration of the fork could be observed by means of a microscope and micrometer eyepiece, and it was found that with care the fork could be made to vibrate with an amplitude which would remain constant within a few per cent. for quite long periods of time.

The microphone was clamped with its axis vertical in a heavy retort-stand, so that it was held at a height of about 1 foot 6 inches above the ground. The grid was connected by a long pair of leads to a Wheatstone's Bridge, which was set up inside a laboratory. A reflecting galvanometer was used, and a preliminary experiment showed that for the small changes in resistance to be observed (not exceeding 0·25 ohm), the deflection shown by the galvanometer was proportional to the resistance change in the microphone.

The fork having been set in vibration, the stand carrying the microphone was placed at a convenient distance and the reading of the galvanometer noted. A piece of card was then placed over the mouth of the glass-bottle resonator, so that the sound from the fork became negligible. This enables the observer to obtain a zero reading on the galvanometer, the difference between the two readings being the deflection due to the sound. The microphone is then moved into another position at a greater or less distance from the fork and the process repeated. The result of one experiment is given below. The distances vary from 12 to 64 feet, and the deflections shown in the table are the means of three or four observations in each position. The actual readings for any particular distance did not differ amongst themselves by more than 0·3 cm., except in the case of the first two deflections, when the maximum variation was 0·5 cm.

---

* The experiment was carried out on Woolwich Common, about one hour after sunset on a calm evening.

| Distance in feet. | Deflection in centimetres. | Distance in feet. | Deflection in centimetres. |
|---|---|---|---|
| 12 | 13·0 | 36 | 1·7 |
| 16 | 9·1 | 40 | 1·4 |
| 20 | 5·7 | 44 | 1·1 |
| 24 | 3·3 | 48 | 0·85 |
| 28 | 3·0 | 56 | 0·8 |
| 32 | 2·3 | 64 | 0·6 |

In fig. 11 the deflection $d$ is plotted against the distance $r$. By plotting log $d$ against log $r$ we obtain a straight line represented by

$$\log_{10} d = 3{\cdot}34 - 2 \log_{10} r,$$

which shows that the deflection $d$ is proportional to $\frac{1}{r^2}$, that is, to the intensity of the sound.

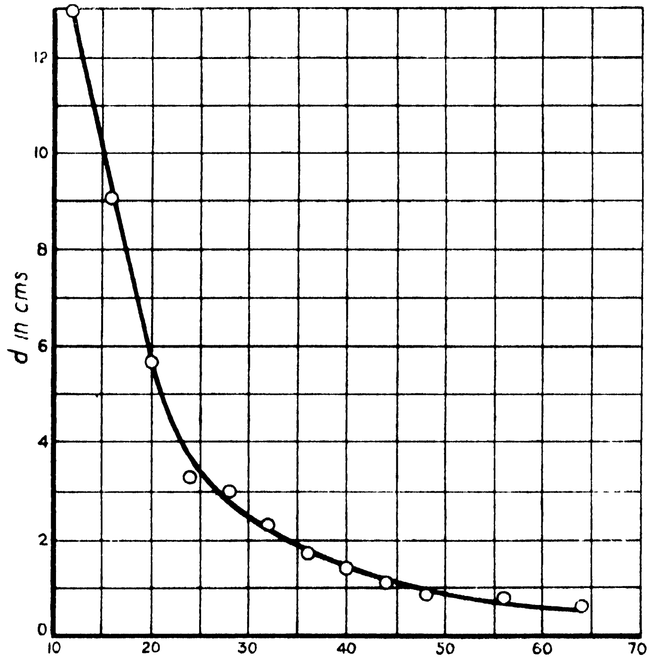


Fig. 11.

According to the equations given in § 6, we should have

$$\delta R_1 = -0{\cdot}23U^2(1-0{\cdot}0072U^2)$$

for all values of U up to about 4 cms. per second. It becomes of interest to estimate what is the probable maximum value of U in the above experiment. It was found that the sound from the fork was certainly inaudible to the unaided ear at a distance of 250 feet, this being probably rather an over-estimate of the distance at which the sound was just lost. Taking the amplitude of the sound at this distance to be $1 \cdot 27 \times 10^{-7}$ cms. (RAYLEIGH's value for the minimum amplitude audible when the note is 256), we find that the amplitude of the sound at 12 feet from the fork is about $2 \cdot 7 \times 10^{-6}$, so that the maximum value of velocity in the sound-wave does not exceed 0·005 cm. per second, and the maximum velocity U of the air in the neck of the resonator will probably not be greater than 0·3 cm. In this case the effect of the second term in the expression for $\delta R_1$ is negligible.

A second experiment performed on another occasion under almost identical conditions confirmed the results already obtained. The distances employed in this experiment, however, did not exceed 32 feet.

(ii) *Second Experiment.*—The expression deduced in § 6 for the oscillatory resistance-change of the same frequency as that of the sound is

$$\delta R_2 = a'U\left(1 + \tfrac{3}{4}\frac{c'}{a'}U^2\right)\sin pt,$$

where $\frac{3}{4}\frac{c'}{a'}$ is very small, so that, except for exceedingly loud sounds,

$$\delta R_2 = a'U \sin pt.$$

In the case of experiments such as that just described, we may write

$$\delta R_2 = \frac{A}{r}\sin pt,$$

since $\delta R_2$ is proportional to the amplitude of the sound affecting the resonator.

When an amplifier is used as a means of observing this effect, its working depends in the first place on the fluctuation of the current in the primary circuit. The effect of the small oscillatory resistance change in the microphone is to produce a " ripple " on the steady heating current, the amplitude of which to a first approximation is proportional to the amplitude of the oscillatory resistance change. Without considering the processes by which this ripple is amplified by the series of transformers and thermionic valves which constitute a transformer amplifier, we shall at once proceed to enquire by an experimental method whether the amplitude of the current on the output side of the amplifier is proportional to the amplitude of the sound affecting the microphone. It is perhaps scarcely necessary to point out that in such an experiment it cannot be assumed that the amplification is constant for different values of the amplitude of the ripple. For ripples of very small amplitude, however, it seems probable that the amplification may be sensibly constant over a moderate range. In spite of these difficulties, which make the interpretation of the observations somewhat obscure, the results obtained appear to be of sufficient interest to justify their inclusion in this paper.

The general procedure followed in the experiment was similar to that already described. The Wheatstone's Bridge was replaced by an Amplifier of the Army pattern known as "C Mark II.," the output terminals being connected to a Campbell vibration-galvanometer tuned to respond to 250 vibrations per second. The source of sound was the same as before. It was not necessary in this case to take a zero reading for each position of the microphone.

The results of two experiments, carried out on different occasions and with different strengths of the source, are given in the following table:—

| Experiment A. | | Experiment B. | |
|---|---|---|---|
| Distance in feet. | Deflection in centimetres. | Distance in feet. | Deflection in centimetres. |
| 8 | 11·4 | 20 | 10·1 |
| 12 | 9·0 | 24 | 7·3 |
| 16 | 7·2 | 28 | 6·1 |
| 20 | 6·0 | 32 | 5·6 |
| 24 | 5·4 | 36 | 4·8 |
| 28 | 5·1 | 40 | 4·4 |
| 32 | 4·6 | 44 | 4·2 |
| 36 | 4·2 | 48 | 4·1 |
| 40 | 3·8 | 52 | 3·9 |
| 44 | 3·2 | 56 | 3·7 |
| 48 | 3·0 | 60 | 3·0 |
| | | 64 | 2·7 |

The curves A and B (fig. 12) show the results of plotting the observations in

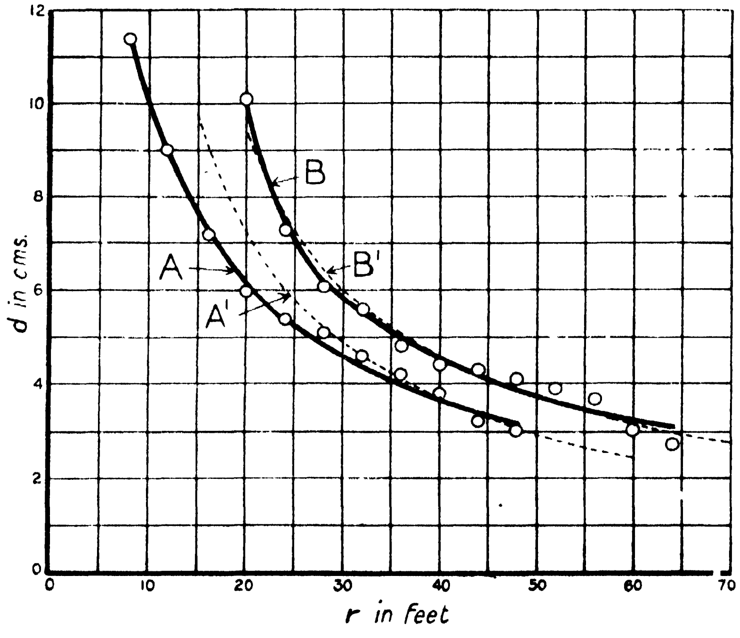


Fig. 12.

experiments A and B respectively. The curve A′ (fig. 12) is calculated from $d = \frac{146 \cdot 3}{r}$. It will be seen that the observed curve A lies along A′ when $r$ is greater than about 28 feet, that is, when the sound has fallen below a certain amplitude. When $r$ is less than 28 feet, the amount of deflection is less than it would have been had the simple inverse first-power rule continued to hold. This is indeed what might be expected from a consideration of the action of the microphone grid, as pointed out at the end of § 6, but the observed falling off in deflection is too great to be accounted for in this way alone.

In experiment B (curve B, fig. 12), the agreement between the observations and the relation $d \propto \frac{1}{r}$ is better. The broken curve B′ is calculated from $d = \frac{185 \cdot 4}{r}$.

Although the deflection corresponding to a given value of $r$ is greater in the B curve than the A curve, it must not be inferred that the sound was louder in the former case. The deflection obtained depends not only on the loudness of the sound but also on the sensitivity to which the amplifier is adjusted and which can be varied over a very wide range. As a matter of fact the sound was louder in the A than in the B experiment. It may also be noted that in both cases the sound was louder than in the Wheatstone's Bridge experiment.

To sum it up, it may be said that the available evidence points to the fact that, when the microphone is employed with an amplifier and vibration galvanometer and only very faint sounds are observed, the deflection shown by the galvanometer is approximately proportional to the amplitude of the sound.

### § 8. *The Effect of Tilting the Microphone.*

It has already been mentioned that, when taking sound measurements with a Hot-Wire Microphone, it is necessary to keep the axis of the microphone inclined at some fixed angle to the horizontal. In fig. 13 a curve has been given showing the relation between the resistance of the grid and the angle of inclination of the axis to the vertical, the electric current carried by the grid being maintained at a constant value throughout the measurements. It is obvious from this curve that the tilt of the axis must not be altered while measurements are being taken by the Wheatstone's Bridge method.

The two most noticeable features about the curve are:

(i) That the resistance of the grid is less when the microphone is held upside down ($\theta = \pi$) than when it is in its normal upright position ($\theta = 0$), although in both cases the plane in which the grid lies is approximately horizontal; and

(ii) That the resistance is least—that is, the rate of cooling is greatest—when $\theta$ is somewhat greater than $\frac{\pi}{2}$.

These two experimental results are curious and difficult to explain. They are both, however, quite characteristic of the form of microphone under discussion, and cannot be attributed entirely to adventitious circumstances, such as the bending of the wire loops of the grid out of their normal positions.

Another point which may be noted is that, since the disposition of the loops about the axis of the microphone is not symmetrical, a rotation of the microphone about its own axis produces a change in resistance. It follows from this that when observations such as those shown in fig. 13 are continued for values of $\theta$ between 0 and $-\pi$, the curve is not in general symmetrical about the line $\theta = 0$.

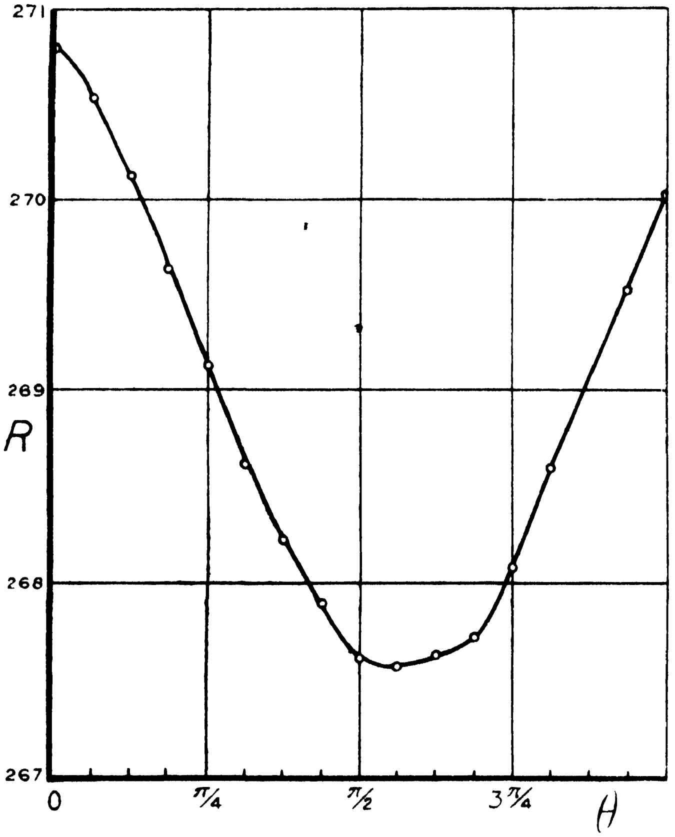


Fig. 13.

It is clear from these results that the resistance of the grid depends, not only on the magnitude of the current of air by which it is cooled, but also on the direction of this current relatively to the plane in which the grid lies. This is due to the free convection from one part of the grid acting upon another part, and is essentially the same phenomenon as that used by J. S. G. THOMAS in the construction of a " Hot-Wire Inclinometer."*

When the Amplifier method is used it is found, as stated previously, that the effect of tilting the microphone is to change the character of the sound heard in the telephones. As $\theta$ is increased from 0 to $\frac{\pi}{2}$ the fundamental note becomes gradually weaker, while at the same time the octave becomes more and more prominent. For some value of $\theta$ exceeding $\frac{\pi}{2}$, but varying somewhat with different grids and the way in which the microphone is held, the fundamental is almost suppressed and the octave is heard with corresponding clearness. When $\theta$ is still further increased the fundamental becomes gradually restored, and the sound of the octave is altogether lost to the ear when $\theta = \pi$. The value of $\theta$ at which the octave is most clearly heard (generally between 110° and 120°) is altered slightly if the microphone is rotated about its own axis.

---

* " An Electrical Hot-Wire Inclinometer," by J. S. G. THOMAS, 'Proc. Phys. Soc.'. Lond., vol. XXXII., pp. 291-314 (1920).

To demonstrate this effect, a microphone (M) was mounted with its neck projecting into a tube (TT) (fig. 14), which could be rotated about a horizontal axis. When so rotated the axis of the microphone could be inclined to the vertical at any desired angle, while the open ends of the tube (TT) were exposed to the same amount of sound throughout the experiment. The microphone was connected to an amplifier in the usual manner, and the output terminals of the amplifier were joined to a vibration galvanometer. The source of sound used was an electrically maintained tuning-fork making 250 vibrations per second, the microphone and vibration galvanometer being also tuned to this frequency.

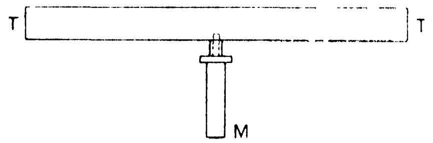


Fig. 14.

The curve in fig. 15 shows the deflection of the vibration galvanometer plotted

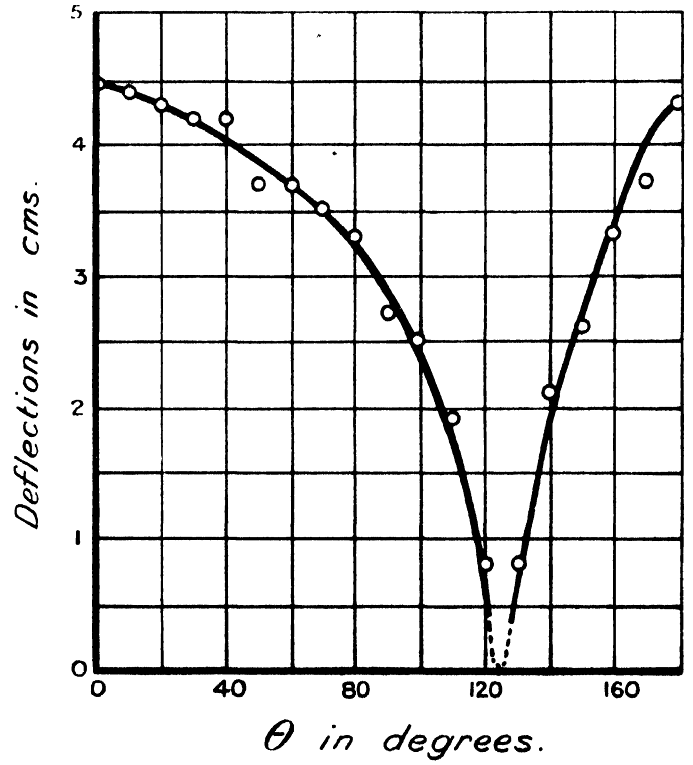


Fig. 15.

against $\theta$, and demonstrates clearly that the amplitude of the fundamental vibration passes through a minimum when $\theta$ is about 125°.

An application of the experimental method of investigation described in § 6 leads to some interesting results. For example, an experiment was performed to determine the relation between the resistance R of the grid and the impressed velocity U of the air-current when $\theta = \frac{\pi}{2}$. The results are given in the following table. The value of $\theta$ indicates only the tilt of the axis of the container and neck, and does not necessarily mean that the plane containing the loops is exactly vertical.

MICROPHONE GRID A1079. Heating current 28·5 milliamperes. Resistance of grid when impressed air-current is zero = 266·85 ohms.

| U. cm. per sec. | δR. ohms. | U. cm. per sec. | δR. ohms. |
|---|---|---|---|
| −5·54 | −9·25 | +1·04 | +0·002 |
| −4·22 | −6·21 | +1·11 | −0·02 |
| −4·15 | 5·64 | +1·58 | 0·31 |
| −2·98 | −2·92 | +2·15 | −0·84 |
| −2·32 | −1·78 | +3·48 | −2·99 |
| −1·36 | 0·93 | +4·49 | −5·46 |
| −1·17 | −0·70 | +5·11 | −6·51 |
| −0·64 | −0·34 | | |

These observations are shown graphically in fig. 16.

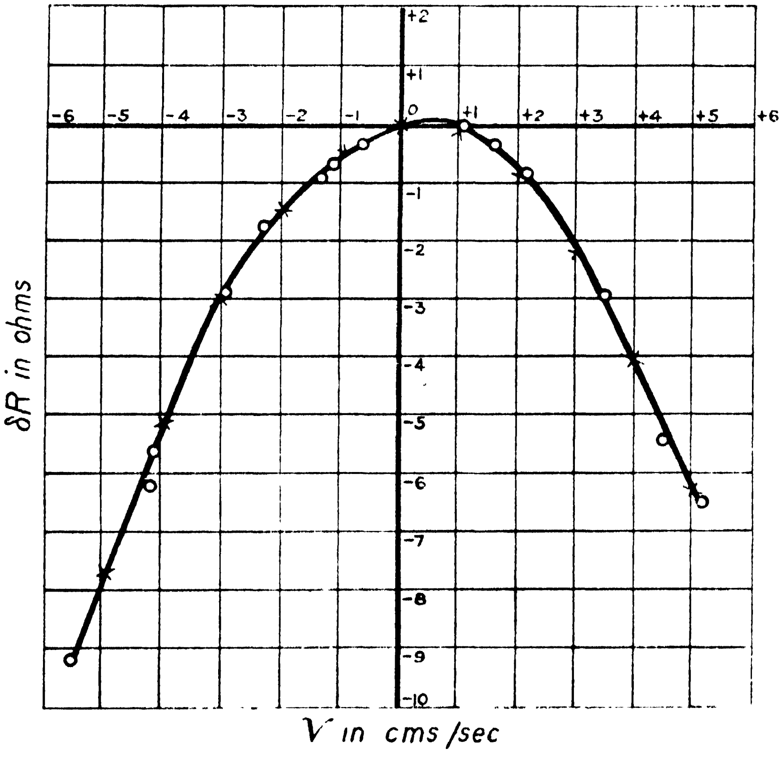


Fig. 16.

Assuming, as in the case when the microphone axis is vertical, that the relation between U and $\delta R$ can be put in the form

$$\delta R = \delta R_0 + a(U-V_0)^2 + b(U-V_0)^4,$$

we find that a fair agreement is obtained when

$$\begin{aligned} \delta R_0 &= +0\cdot 02 \\ a &= -0\cdot 3 \\ b &= +0\cdot 0007 \\ V_0 &= +0\cdot 25 \end{aligned}$$

that is,

$$\delta R = 0\cdot 02 - 0\cdot 3(U-0\cdot 25)^2 + 0\cdot 0007(U-0\cdot 25)^4.$$

Points calculated from this formula when $U = 0, \pm 1$, etc., are shown by the crosses in fig. 16. It will be seen that the fit is not so good as in the case when the microphone was held vertically (fig. 10), and also that the experimental curve is not quite symmetrical about the line $U = V_0 = 0\cdot 25$ when U is small. It is, however, difficult to obtain reliable readings of the resistance of the grid when the microphone axis is horizontal or nearly so, and the above formula appears to represent the experimental results sufficiently well for the present purpose. It may be written

$$\delta R = +0\cdot 15U - 0\cdot 3U^2 - 0\cdot 00035U^3 + 0\cdot 0007U^4,$$

and by putting U sin *pt* in place of U, and disregarding the terms in $U^3$ and $U^4$, it can be seen that the three principal effects to be expected when the microphone is exposed to a sound-wave of suitable frequency are

$$\begin{aligned} \delta R_1 &= -0\cdot 15U^2, \\ \delta R_2 &= +0\cdot 15U \sin pt, \\ \delta R_3 &= +0\cdot 15U^2 \cos 2pt. \end{aligned}$$

If these are compared with the corresponding expressions in § 6, it will be seen that for a given value of U the magnitude of the steady resistance change and the amplitude of the octave are reduced in the proportion 15/23, but that the amplitude of the fundamental vibration is reduced to 15/119 of its value when $\theta = 0$. Although, therefore, all these effects are diminished when the microphone is laid horizontally, it is the fundamental vibration which suffers the most. So far the deductions made from the results of the experiments with steady air-currents are in accordance with observation.

For values of $\theta$ between 0 and $\frac{\pi}{2}$, curves intermediate in form between those in figs. 10 and 16 are obtained. When $\theta$ is greater than 0, $V_0$ no longer represents the velocity of free convection from the grid, but is more nearly the component of the free convection along the axis of the microphone. That is, $V_0$ is approximately equal to $\bar{V}\cos\theta$, $\bar{V}$ being the velocity of free convection from the grid. If this were strictly true, we should have $V_0 = 0$ when $\theta = \frac{\pi}{2}$, but this is not generally the case, probably owing to the fact that the loops are slightly displaced and do not lie in a plane which is just at right angles to the microphone axis, and it may even happen that they are not all in the same plane. For example, in the case of the grid used in the experiment just described, it was found that

$$V_0 = 1\cdot225 \text{ cms. per second when } \theta = 0,$$

and

$$V_0 = 0\cdot25 \text{ cm. per second when } \theta = \frac{\pi}{2}.$$

As an approximation we may take $\bar{V} = 1\cdot225$ and $V_0 = \bar{V}\cos(\theta+\alpha)$ when $\theta$ is near $\frac{\pi}{2}$, so that

$$1\cdot225\cos\left(\frac{\pi}{2}+\alpha\right) = 0\cdot25.$$

Whence

$$\alpha = -11^\circ,$$

and we should have $V_0 = 0$ when $\theta = 101^\circ$.

When $V_0 = 0$ the expressions for the resistance changes produced by an oscillatory air-current (§ 6) reduce to

$$\delta R_1 = \tfrac{1}{2}aU^2,$$
$$\delta R_2 = 0,$$
$$\delta R_3 = -\tfrac{1}{2}aU^2\cos 2pt,$$

so that the total resistance change is made up of two parts only—a steady change and the octave—while the fundamental vibration is completely suppressed. Now the curve in fig. 15 shows that in that particular experiment the fundamental was suppressed when $\theta$ was about 125 degrees, and this is indeed about the usual value of $\theta$ for this phenomenon to occur, while (as in the above example) $V_0$ vanishes when $\theta$ is at most about 100 degrees. It appears therefore, that merely writing $U\sin pt$ instead of $U$, in the equation connecting $U$ and $\delta R$ for small steady velocities, does not in this case explain all the observed phenomena.

A satisfactory explanation of this difficulty has not yet been found, but the hypothesis that a jet may be formed in the neck of the resonator may be put forward. The possibility of this occurring in the mouths of ordinary resonators is discussed by RAYLEIGH ("Theory of Sound," vol. II., § 322). If a jet were formed in the neck of the resonator, then, in order to account for the observed phenomenon, it would have to be of such a nature that the outward movement of the air takes place principally along the central axis of the neck, while the inward movement takes place close to the walls of the neck. Since the platinum grid is placed centrally in the neck and does not occupy any space near the walls, the effect of the jet could be represented by adding a term $V'$ to $V_0$, so that (approximately)

$$V_0 = \bar{V}\cos(\theta+\alpha)+V'.$$

The fundamental would then be suppressed when

$$\bar{V}\cos(\theta+\alpha)+V' = 0.$$

If $\alpha = -11°$ (as in the case of the grid used in the experiment described above), then $\theta$ will be 125° if $\frac{V'}{\bar{V}} = 0·4$.

A further point in favour of this hypothesis is that the angle at which the microphone must be tilted for the octave to be heard best depends to some extent on the intensity of the sound.

On the other hand, if a jet were formed, then since the amplitude of the fundamental is approximately proportional to $V_0$, the amplitude when $\theta = 0$ should be to the amplitude when $\theta = \pi$ in the approximate ratio $\frac{\bar{V} + V'}{\bar{V} - V'}$. And since we must have $\frac{V'}{\bar{V}}$ about equal to 0·4 in order to make $\theta = 125$ degrees when the octave is heard best, it follows that the microphone should be more than twice as sensitive to the fundamental vibration when it is held upright than when it is held upside down. This is not borne out by experiment, which shows that the sensitivity is only slightly reduced by turning the microphone upside down (see fig. 15).

RAYLEIGH* points to the near agreement between observed and calculated pitch in support of his view that jets are not formed " to any appreciable extent at the mouths of resonators as ordinarily used." The further argument, however, that " the persistence of the free vibration . . . seems to exclude any important cause of dissipation beyond the communication of motion to the surrounding air," does not apply to the resonators used in the present experiments, for it is shown in § 4 that the dissipation caused by the communication of motion to the surrounding air is negligible compared with the total amount of dissipation which occurs.

It appears, therefore, that the jet hypothesis, while offering a plausible explanation of the suppression of the fundamental at such large values of $\theta$ as 125 degrees, is open to objection on account of

(i) The nearly equal sensitivity shown by the microphone in the erect and inverted positions ; and

(ii) The near agreement between observed and calculated pitch (see § 4).

---

* *Loc. cit.*, p. 217.

### § 9. *Some Observations of Distribution and Intensity of Sound made with the Hot-Wire Microphone.*

The applications of the microphone to sound measurements are sufficiently numerous to justify a short description. It is quite obvious that the apparatus is not adapted to the measurement of the total quantity of a medley of sounds, since the microphones are selective, and the pitch or wave-length of the sound measured must therefore always be given.

Full advantage can be taken of the two alternative methods of using the microphone for sound measurement. In general, it should be laid down that the Wheatstone's Bridge method should be adopted for cases in which the sound distribution can be altered by keeping the microphone fixed in space and changing the position of the source of sound or by movement of any screen, reflector, trumpet, etc., while source and microphone occupy given positions.

If, however, the microphone has to be attached to some moving object the amplifier method has to be employed, records of amplitude being given by a vibration galvanometer, or in certain cases by rectifying the current and using a reflecting galvanometer.

The simplest experiment to perform is that of observing the distribution of intensity of sound in a closed room. By the Wheatstone's Bridge method the microphone can occupy some fixed position while a steady source of sound such as a tuning-fork is carried about the room.

The effects are sufficiently strong for a pivoted galvanometer to be employed for observation.

A variant of this experiment is to keep the source of sound in one place, and observe the effects of moving either one's self, of altering the position of furniture in the room, or of opening of doors or windows. It is quite easy to vary these arrangements in such a manner as to reduce the intensity of sound as recorded by the microphone from a maximum value to zero. The position of nodes and antinodes in the room can be investigated by moving the microphone and employing the amplifier method with telephones or vibration galvanometer. The results obtained are sufficiently striking to condemn any method of sound measurement in a closed room—the mere movement of the observer being sufficient to vitiate any experimental results. Methods of ear testing, which so commonly employ tuning-forks, are equally unsatisfactory. All measurements must therefore be made in the open-air and full precautions must be taken to avoid obstacles presenting reflecting surfaces.

Moreover, open-air work can only be performed under exceptionally calm conditions such as exist on certain nights or during a fog.

These phenomena are of far reaching importance in architectural acoustics. It seems evident that in any room or concert hall there is a considerable difference between the musical piece as rendered by instruments and the sounds which the audience observes; and it also follows that various members of the audience hear the same rendering somewhat differently owing to their positions in the room. That the sensation of suppression of a note is not obvious is no doubt due in some measure to the distance between the two ears.

*Experiments on Reflection.*—A large sheet of uralite was set up out of doors with its plane vertical. An electrically driven tuning-fork served as the source of sound, and was mounted opposite the centre of the uralite at a distance of 30 feet from it. The sound was thus incident normal to the surface and capable of producing stationary waves.

The curve connecting deflection of the galvanometer and distance from the reflector is shown in fig. 17 and indicates clearly the position of nodes and antinodes. The distance between the nodes agrees well with that obtained from a wave-length of 43 inches, except for that nearest to the mirror. All observations taken with various surfaces showed that the first antinode was nearer the mirror surface than was anticipated, which may be due to the lack of rigidity of the reflecting surface. The effect at the reflecting surface varies very largely with its nature; thus, when a wooden door is employed, the effect is greatest at the centre of the door midway between the four panels and least at the panels where there is minimum rigidity.

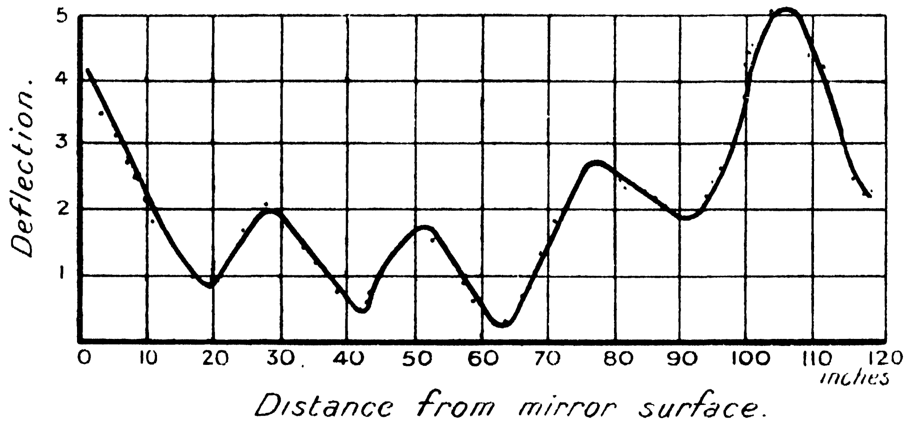


Fig. 17.

The reflecting qualities of different surfaces for sound can thus be compared. It is also obviously a simple matter to test the transmitting properties of various media—taking care to confine the sound transmitted to the material under test.

*Experiments with Trumpets.*—A trumpet has certain magnifying and directional properties, which depend on its dimensions and the wave-length of the sound employed; and another important factor in magnification is the material of which the trumpet is made. In the experiments described below the trumpet employed was conical, having a mouth 18 inches in diameter, a throat ½ inch in diameter and a slant side of 25½ inches. It was made of 1-inch wood in 16 segments.

The trumpet was mounted on a stand with its axis horizontal, and was capable of rotating about a vertical axis, its bearing being indicated by a pointer travelling over a horizontal circle graduated in degrees. The narrow end of the trumpet received the aperture of a resonant microphone, and the connection with the trumpet was such as to leave unchanged the resonant note of the microphone by modification of its orifice.

The source of sound was the electrically maintained tuning-fork previously referred to and its distance from the trumpet was 50 yards. In order to enhance the source of sound the prongs of the tuning-fork were caused to set in resonant vibration the air in a glass bottle, and the mouth of the bottle was taken as the position of the source.

To get the zero of bearing, cross-threads were fixed to the trumpet mouth so that they coincided with two perpendicular diameters of the mouth. A small sighting-hole was drilled in a brass plate covering the narrow end of the trumpet. By looking through this aperture and rotating the trumpet until the cross-threads appeared in line with the source one was able to observe the zero of bearing on the scale.

The following table gives readings of the vibration galvanometer for various orientations of the trumpet :—

| Bearing in degrees. | Deflection in divisions. | Bearing in degrees. | Deflection in divisions. |
|---|---|---|---|
| 0 | 48 | 50 | 36 |
| 5 | 48 | 55 | 34 |
| 10 | 47·5 | 60 | 31 |
| 15 | 47 | 65 | 28 |
| 20 | 46 | 70 | 25 |
| 25 | 45 | 75 | 22 |
| 30 | 43·5 | 80 | 18·5 |
| 35 | 42 | 85 | 15 |
| 40 | 40 | 90 | 11 |
| 45 | 38 | | |

It is interesting to note that the intensity of sound at the throat of the trumpet increases again as the bearing approaches 180 degrees and gives a maximum. Such an effect can be easily observed by fitting a stethoscope to the narrow end of the trumpet and listening by ear as the trumpet is rotated.

When the trumpet is removed and the sound recorded by a microphone alone, an estimate of magnification is given. It was found that the ratio

$$\frac{\text{Maximum deflection with the trumpet}}{\text{Deflection without the trumpet}} = 14 \cdot 5,$$

which is a measure of amplitude magnification, since it has been shown in § 7 that deflection is proportional to amplitude.

The diagram (fig. 18) shows the nature of the polar curve of amplitudes. Experiments with sources of different pitches indicated that the higher the pitch the sharper the curvature in the region of the zero bearing. For highly directive apparatus, therefore, every advantage is to be gained by using big trumpets.

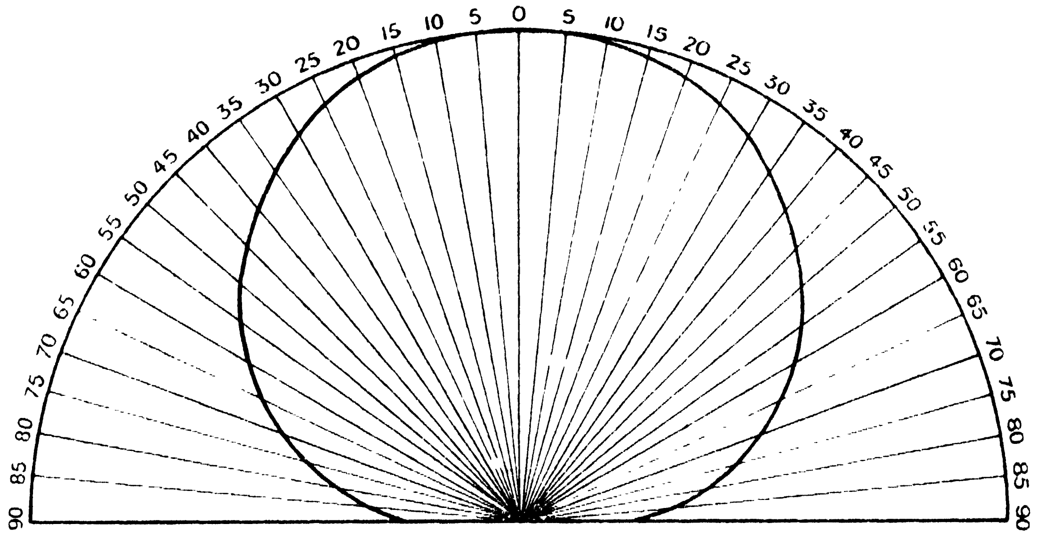


Fig. 18.

The Hot-Wire Microphone gives an easy method of testing the resonance frequencies of trumpets, but the effect must not be complicated by using a resonant microphone of the type described above. In this case the trumpet itself may be used as the resonating cavity. A bare grid with an orifice of the type above described is fitted to the narrow end of the trumpet, as shown in fig. 19, and the amplifier is used with rectifier and

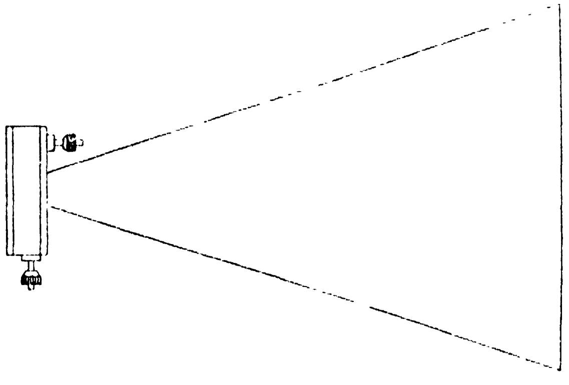


Fig. 19.

reflecting galvanometer. The bridge method cannot be employed in this case as the open trumpet is subject to draughts, and there is constantly a movement of air in one direction or the other which would cause a change in resistance of the heated grid.

The source of sound is the specially constructed siren referred to in § 4. As the pitch rises a deflection is produced in the neighbourhood of resonance. The diagram (fig. 20) shows a relation between the pitch of the siren note and the response of the trumpet. The amount of resonance is expressed in terms of galvanometer deflection as shown by the following table:—

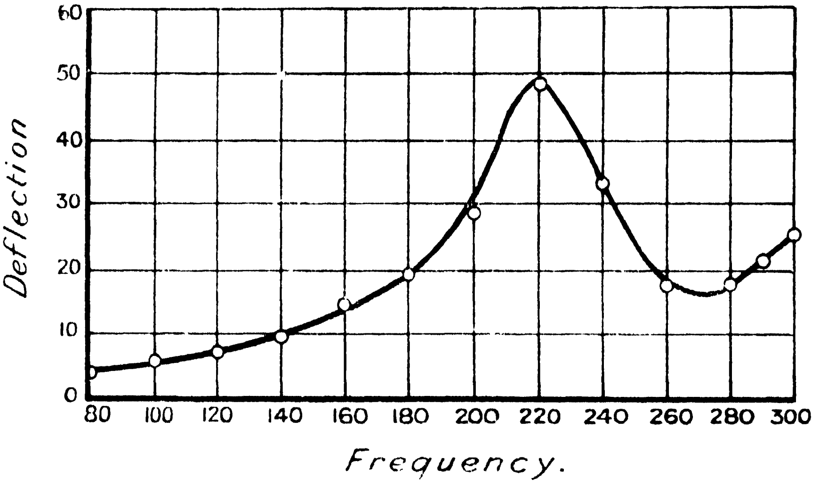


Fig. 20.

| Frequency (vibrations per sec.). | Deflection (divisions). | Frequency (vibrations per sec.). | Deflection (divisions). |
|---|---|---|---|
| 80 | 3 | 200 | 30 |
| 100 | 5·5 | 220 | 49 |
| 120 | 7·5 | 240 | 32 |
| 140 | 10 | 260 | 17 |
| 160 | 13 | 280 | 16 |
| 180 | 20 | 300 | 24 |

The table shows an upward curve at the highest frequency, thus indicating approach to the next overtone. The maximum at 220 indicates the fundamental resonance note.

By means of the microphone, therefore, the whole properties of a trumpet as a receiver of sound can be investigated, both as regards directive action and resonance.

Since by the principle of reversibility we may employ the trumpet as a transmitter for any given note, we may also derive its properties as a distributor and magnifier of any sound produced at the narrow end.

This has an obvious application to gramophone trumpets, in which the diaphragm acts as the source of sound.

In conclusion, one example may be given of the use of the microphone to measure diffraction of sound.

An interesting example is that of the diffraction effect of a single disc. A large wooden disc, 1 inch thick and 10 feet in diameter, is suspended by one edge. The tuning-fork described above serves as a steady source of sound and is placed opposite the centre of the disc and 30 feet from it. The microphone, in tune with this fork, is mounted with its orifice at the centre of the back of the disc—the axis of the microphone being of course vertical. The disc, with microphone, is now swung round about its vertical diameter and readings are taken with the vibration galvanometer—using the Amplifier method. The bearings were observed by means of a sighting-tube passing normally through the disc at a point on its vertical diameter about one-quarter of the way from its lower edge. As the disc was rotated one could observe through the sighting-tube a number of white posts, which were driven into the ground on the circumference of a circle of 50 yards radius, the sighting-tube being at the centre. The pegs were 10 degrees apart, and the zero reading was given when the central post, the source of sound, and the sighting tube were in line.

| Bearing (degrees). | Deflection (divisions). | Bearing (degrees). | Deflection (divisions). |
|---|---|---|---|
| 90 | 45 | 0 | 43 |
| 80 | 31·5 | −10 | 16 |
| 70 | 17 | −20 | Min. 2 |
| 60 | 5·5 | −30 | 10 |
| 52 | Min. 1·4 | −37 | Max. 14·5 |
| 50 | 1·5 | −40 | 12 |
| 40 | 12·5 | −50 | 2 |
| 37 | Max. 14·5 | −51 | Min. 2 |
| 30 | 10 | −60 | 6 |
| 20 | 2 | −70 | 18·5 |
| 19 | Min. 2·0 | | |
| 10 | 20 | −80 | 32 |
| 0 | 43 | −90 | 45 |

If we plot bearing and deflection, which for faint sounds measures the amplitude of the vibration, a curve of the form shown in fig. 21 is obtained.

It is thus seen that the diffraction gives a central maximum equal in intensity to the sound which passes the edge of the disc, and this is surrounded by a ring maximum.

A variant of this experiment was performed in which the disc was set to give different angles of incidence for the sound, and the microphone was moved along a horizontal diameter until a maximum effect was given. One thus obtains an image of the source for each angle of incidence, and the distance of this image from the centre of the disc gradually increases as the angle of incidence is increased.

The image, however, becomes ill-defined for angles of incidence exceeding 20 degrees, and the tendency is then to obtain a train of maxima of nearly equal intensity when the microphone moves from one edge of the disc to the other.

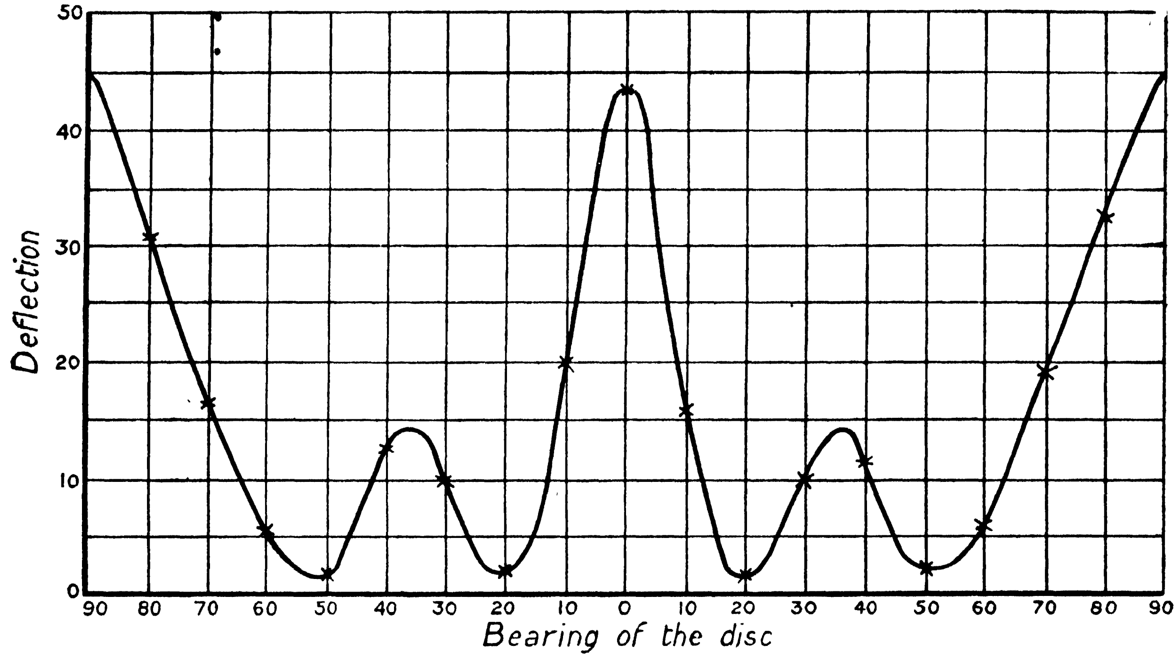


Fig. 21.

The above illustrations give some indication of the manner in which the microphone can be applied to the investigation of acoustical problems. Many of the measurements described in this paper were made in a locality where a certain amount of noise was constantly occurring, but which the microphone, being highly tuned, failed to record. Tuned reception for sound has all the advantages of tuned reception in "wireless" in distinguishing and magnifying faint signals.

A distinct limitation of this microphone is its restriction to the measurement of low-frequency sounds, but it is hoped to devise a microphone of the Hot-Wire type sufficiently sensitive to deal with speech frequencies.

## § 10. *Summary.*

A new form of Selective Hot-Wire Microphone is described, consisting of an electrically heated grid of fine platinum wire placed in the neck of a Helmholtz resonator. The effect of a sound having the same frequency as that natural to the resonator itself is to produce an oscillatory motion of the air in the neck, which in turn causes a change in resistance of the platinum wire grid. The total resistance change comprises a steady fall in resistance due to an average cooling of the grid, and a periodic change due to th to-and-fro motion of the air. Two methods of using the microphone are described :—

(i) A Bridge method, depending on the steady drop in resistance ; and
(ii) An Amplifier method which makes use of the periodic resistance changes.

Curves are given showing the sharpness of resonance as measured by the Bridge method.

The various factors affecting the sensitivity of the microphone are discussed. The most important, from a practical point of view, is the variation of the sensitivity with the heating current of the grid. It is found by experiment that the sensitivity always increases as the heating current is increased. In the case of the Bridge method, it is found that the steady resistance change produced by a sound of given intensity is a linear function of the temperature of the grid above its surroundings measured on the platinum scale.

The results of experiments on the cooling of the grid by low velocity air-currents are described. From these results it is deduced that the principal resistance changes to be expected when the grid is cooled by an oscillatory air-currents are :—

(1) A steady drop due to an average cooling ;
(2) A periodic resistance change at the same frequency as that of the sound ; and
(3) A periodic resistance change of frequency twice that of the sound.

All these effects are found in practice.

Further deductions are that the steady change of resistance is proportional to the *intensity* of the sound, while the periodic resistance change in (2) is proportional to the *amplitude*. These conclusions are confirmed by experiment.

A description is given of the effect to be observed when the microphone is tilted at various angles, and the observed facts are compared with what would be expected from the results of experiments with steady air-currents.

Finally, an account is given of some experiments which exemplify the use of the Hot-Wire Microphone for observing the intensity and distribution of sound.

The work described in this paper forms part of an investigation commenced in the Munitions Inventions Department, and continued later at the Signals Experimental Establishment, Woolwich.

In conclusion, the authors wish to express their indebtedness to the Chief Experimental Officer of this Establishment for the interest which he has taken in the progress of the work, and for the facilities which they have received for carrying out experiments.

# INDEX

TO THE

# PHILOSOPHICAL TRANSACTIONS.

## SERIES A, VOL. 221.

A.

Aberration, theory of second order longitudinal spherical (Baker and Filon), 29.

Aerodynamics of spinning shell (Fowler and others), 295.

Atmospheric turbulence, measurements of (Richardson), 1.

B.

Baker (T. Y.) and Filon (L. N. G.). On a Theory of the Second Order Longitudinal Spherical Aberration for a Symmetrical Optical System, 29.

C.

Chakko (K. C.). See Coker and Chakko.

Coker (E. G.) and Chakko (K. C.). The Stress-strain Properties of Nitro-cellulose and the Law of its Optical Behaviour, 139.

D.

Dalby (W. E.). Researches on the Elastic Properties and the Plastic Extension of Metals, 117.

E.

Error, reduction by linear compounding (Sheppard), 199.

F.

Filon (L. N. G.). See Baker and Filon.

Fowler (R. H.) (and others). The Aerodynamics of a Spinning Shell, 295.

G.

Gallop (E. G.). See Fowler and others.

Griffith (A. A.). The Phenomena of Rupture and Flow in Solids, 163.

J.

Jeffery (G. B.). Plane Stress and Plane Strain in Bipolar Co-ordinates, 265.

Jeffreys (H.). Tidal Friction in Shallow Seas, 239.

[Published March 16, 1921.

L.

Lightning discharges and electric field of thunderstorms (**Wilson**), 73.
Lock (C. N. H.). See **Fowler** and others.

M.

Metals, elastic properties and plastic extension of (**Dalby**), 117.
Microphone, selective hot-wire (**Tucker** and **Paris**), 389.

N.

Nitro-cellulose, stress-strain properties and optical behaviour (**Coker** and **Chakko**), 139.

O.

Optical system, longitudinal spherical aberration for symmetrical (**Baker** and **Filon**), 29.

P.

Paris (E. T.). See **Tucker** and **Paris**.

R.

Richardson (L. F.). Some Measurements of Atmospheric Turbulence, 1.
Richmond (H. W.). See **Fowler** and others.
Rupture and flow in solids (**Griffith**), 163

S.

Shell, aerodynamics of spinning (**Fowler** and others), 295.
Sheppard (W. F.). Reduction of Error by Linear Compounding, 199.
Stress and strain in bipolar co-ordinates (**Jeffery**), 265.

T.

Thunderstorms, electric field of (**Wilson**), 73.
Tidal friction in shallow seas (**Jeffreys**), 239.
Tucker (W. S.) and Paris (E. T.) A Selective Hot-Wire Microphone, 389

W.

Wilson (C. T. R.). Investigations on Lightning Discharges and on the Electric Field of Thunderstorms, 73.

---

HARRISON AND SONS, LTD., PRINTERS IN ORDINARY TO HIS MAJESTY, ST. MARTIN'S LANE, W.C. 2.

# THE PHILOSOPHICAL TRANSACTIONS OF THE ROYAL SOCIETY

| | £ | s. | d. |
|---|---|---|---|
| 1801. Part II... | 0 | 17 | 6 |
| 1802. Part I... | 0 | 11 | 0 |
| —— Part II... | 0 | 17 | 6 |
| 1803. Part I... | 0 | 12 | 6 |
| —— Part II... | 0 | 13 | 6 |
| 1804. Part I... | 0 | 10 | 6 |
| —— Part II... | 0 | 12 | 6 |
| 1805. Part I... | 0 | 10 | 0 |
| —— Part II... | 0 | 11 | 6 |
| 1806. Part I... | 0 | 13 | 6 |
| —— Part II... | 0 | 17 | 6 |
| 1807. Part I... | 0 | 10 | 0 |
| —— Part II... | 0 | 15 | 6 |
| 1808. Part I... | 0 | 9 | 6 |
| —— Part II... | 0 | 14 | 6 |
| 1809. Part I... | 0 | 14 | 6 |
| —— Part II... | 1 | 0 | 6 |
| 1810. Part I... | 0 | 10 | 6 |
| —— Part II... | 0 | 12 | 0 |
| 1811. Part I... | 0 | 15 | 0 |
| —— Part II... | 0 | 15 | 0 |
| 1812. Part I... | 0 | 17 | 6 |
| —— Part II... | 0 | 17 | 6 |
| 1813. Part I... | 0 | 14 | 0 |
| —— Part II... | 0 | 18 | 0 |
| 1814. Part I... | 0 | 18 | 0 |
| 1815. Part II... | 1 | 2 | 0 |
| 1816. Part II... | 0 | 17 | 6 |
| 1819. Part III... | 0 | 10 | 0 |
| 1823. Part II... | 1 | 8 | 0 |
| 1824. Part I... | 0 | 12 | 6 |
| —— Part II... | 1 | 0 | 0 |
| —— Part III... | 1 | 4 | 0 |
| 1825. Part I... | 1 | 4 | 0 |
| —— Part II... | 2 | 0 | 0 |
| 1826. Part I... | 1 | 2 | 6 |
| —— Part II... | 0 | 12 | 6 |
| —— Part III... | 2 | 0 | 0 |
| —— Part IV... | 1 | 2 | 6 |
| 1827. Part I... | 0 | 18 | 0 |
| —— Part II... | 0 | 18 | 0 |
| 1828. Part I... | 1 | 1 | 0 |
| —— Part II... | 0 | 10 | 0 |
| 1829. Part I... | 0 | 16 | 0 |
| —— Part II... | 0 | 14 | 0 |
| 1830. Part I... | 1 | 10 | 0 |
| —— Part II... | 1 | 1 | 0 |
| 1831. Part I... | 1 | 10 | 0 |
| —— Part II... | 1 | 12 | 0 |
| 1832. Part I... | 1 | 1 | 0 |
| —— Part II... | 2 | 0 | 0 |
| 1833. Part I... | 1 | 1 | 0 |
| —— Part II... | 2 | 18 | 0 |
| 1834. Part I... | 0 | 17 | 0 |
| —— Part II... | 2 | 2 | 0 |

| | £ | s. | a. |
|---|---|---|---|
| 1835. Part I... | 1 | 2 | [illegible] |
| —— Part II... | 0 | 14 | 0 |
| 1836. Part I... | 1 | 10 | 0 |
| —— Part II... | 2 | 0 | 0 |
| 1837. Part I... | 1 | 8 | 0 |
| —— Part II... | 1 | 8 | 0 |
| 1838. Part I... | 0 | 13 | 0 |
| —— Part II... | 1 | 8 | 0 |
| 1839. Part I... | 0 | 18 | 0 |
| —— Part II... | 1 | 1 | 6 |
| 1840. Part I... | 0 | 18 | 0 |
| —— Part II... | 2 | 5 | 0 |
| 1841. Part I... | 0 | 10 | 0 |
| —— Part II... | 1 | 10 | 0 |
| 1842. Part I... | 0 | 16 | 0 |
| —— Part II... | 1 | 2 | 0 |
| 1843. Part I... | 0 | 10 | 0 |
| —— Part II... | 1 | 10 | 0 |
| 1844. Part I... | 0 | 10 | 0 |
| —— Part II... | 1 | 10 | 0 |
| 1845. Part I... | 0 | 16 | 0 |
| —— Part II... | 1 | 0 | 0 |
| 1846. Part I... | 0 | 7 | 6 |
| —— Part II... | 1 | 12 | 0 |
| —— Part III... | 1 | 12 | 0 |
| —— Part IV... | 1 | 12 | 0 |
| 1847. Part I... | 0 | 14 | 0 |
| —— Part II... | 0 | 16 | 0 |
| 1848. Part I... | 1 | 0 | 0 |
| —— Part II... | 0 | 14 | 0 |
| 1849. Part I... | 1 | 0 | 0 |
| —— Part II... | 2 | 5 | 0 |
| 1850. Part I... | 1 | 10 | 0 |
| —— Part II... | 3 | 5 | 0 |
| 1851. Part I... | 2 | 10 | 0 |
| —— Part II... | 2 | 10 | 0 |
| 1852. Part I... | 1 | 0 | 0 |
| —— Part II... | 2 | 5 | 0 |
| 1853. Part I... | 0 | 18 | 0 |
| —— Part II... | 0 | 12 | 0 |
| —— Part III... | 1 | 2 | 0 |
| 1854. Part I... | 0 | 12 | 0 |
| —— Part II... | 0 | 16 | 0 |
| 1855. Part I... | 0 | 16 | 0 |
| —— Part II... | 1 | 6 | 0 |
| 1856. Part I... | 2 | 0 | 0 |
| —— Part II... | 1 | 4 | 0 |
| —— Part III... | 1 | 4 | 0 |
| 1857. Part I... | 1 | 8 | 0 |
| —— Part II... | 1 | 4 | 0 |
| —— Part III... | 1 | 2 | 0 |
| 1858. Part I... | 1 | 8 | 0 |
| —— Part II... | 3 | 0 | 0 |
| 1859. Part I... | 2 | 10 | 0 |
| —— Part II... | 2 | 5 | 0 |
| 1860. Part I... | 0 | 16 | 0 |
| —— Part II... | 2 | 1 | 6 |
| 1861. Part I... | 1 | 3 | 0 |
| —— Part II... | 1 | 5 | 0 |
| —— Part III... | 1 | 7 | 6 |
| 1862. Part I... | 2 | 14 | 0 |
| —— Part [illegible] | 3 | 0 | 0 |
| 1863. Pa[illegible] | [illegible] | 4 | 0 |
| —— Part II... | | | |

| | £ | s. | d. |
|---|---|---|---|
| 1864. Part I... | 0 | 11 | 0 |
| —— Part II... | 1 | 7 | 6 |
| —— Part III... | 1 | 10 | 0 |
| 1865. Part I... | 2 | 2 | 0 |
| —— Part II... | 1 | 5 | 0 |
| 1866. Part I... | 1 | 14 | 0 |
| —— Part II... | 2 | 7 | 6 |
| 1867. Part I... | 1 | 3 | 0 |
| —— Part II... | 1 | 15 | 0 |
| 1868. Part I... | 2 | 5 | 0 |
| —— Part II... | 2 | 0 | 0 |
| 1869. Part I... | 2 | 10 | 0 |
| —— Part II... | 3 | 3 | 0 |
| 1870. Part I... | 1 | 10 | 0 |
| —— Part II... | 1 | 18 | 0 |
| 1871. Part I... | 1 | 10 | 0 |
| —— Part II... | 2 | 5 | 0 |
| 1872. Part I... | 1 | 12 | 0 |
| —— Part II... | 2 | 8 | 0 |
| 1873. Part I... | 2 | 10 | 0 |
| —— Part II... | 1 | 5 | 0 |
| 1874. Part I... | 2 | 8 | 0 |
| —— Part II... | 3 | 0 | 0 |
| 1875. Part I... | 3 | 0 | 0 |
| —— Part II... | 3 | 0 | 0 |
| 1876. Part I... | 2 | 8 | 0 |
| —— Part II... | 2 | 8 | 0 |
| 1877. Part I... | 1 | 16 | 0 |
| —— Part II... | 2 | 5 | 0 |
| Vol. 168 (extra) | 3 | 0 | 0 |
| 1878. Part I... | 1 | 16 | 0 |
| —— Part II... | 3 | 0 | 0 |
| 1879. Part I... | 2 | 0 | 0 |
| —— Part II... | 1 | 12 | 0 |
| 1880. Part I... | 2 | 5 | 0 |
| —— Part II... | 2 | 0 | 0 |
| —— Part III... | 1 | 1 | 0 |
| 1881. Part I... | 2 | 10 | 0 |
| —— Part II... | 1 | 10 | 0 |
| —— Part III... | 2 | 2 | 0 |
| 1882. Part I... | 1 | 14 | 0 |
| —— Part II... | 2 | 0 | 0 |
| —— Part III... | 2 | 10 | 0 |
| —— Part IV... | 1 | 0 | 0 |
| 1883. Part I... | 1 | 10 | 0 |
| —— Part II... | 2 | 10 | 0 |
| —— Part III... | 1 | 12 | 0 |
| 1884. Part I... | 1 | 8 | 0 |
| —— Part II... | 1 | 16 | 0 |
| 1885. Part I... | 2 | 10 | 0 |
| —— Part II... | 2 | 5 | 0 |
| 1886. Part I... | 1 | 8 | 0 |
| —— Part II.... | 1 | 15 | 0 |
| 1887. (A.) .... | 1 | 3 | 0 |
| —— (B.) .... | 1 | 16 | 0 |
| 1888. (A.) .... | 1 | 10 | 0 |
| —— (B.) .... | 2 | 17 | 6 |
| 1889. (A.) .... | 1 | 18 | 0 |
| —— (B.) .... | 1 | 14 | 0 |
| 1890. (A.) .... | 1 | 16 | 6 |
| —— (B.) .... | 1 | 5 | 0 |
| 1891. (A.) .... | 2 | 2 | 0 |
| —— (B.) .... | 3 | 3 | 0 |
| 1892. (A.) .... | 2 | 1 | 0 |
| —— (B.) .... | 2 | 2 | 0 |
| 1893. (A.) .... | 3 | 14 | 0 |
| —— (B.) .... | 2 | 13 | 0 |

| | £ | s. | d |
|---|---|---|---|
| 1894. (A.) Pt. I. | 1 | 5 | 0 |
| —— (A.) Pt. II. | 1 | 10 | 0 |
| —— (B.) Pt. I. | 3 | 10 | 0 |
| —— (B.) Pt. II. | 1 | 11 | 6 |
| 1895. (A.) Pt. I. | 1 | 14 | 0 |
| —— (A.) Pt. II. | 1 | 3 | 6 |
| —— (B.) Pt. I. | 1 | 15 | [illegible] |
| —— (B.) Pt. II. | 0 | 16 | 6 |
| 1896 (A) Vol. 187. | 1 | 19 | 6 |
| —— (A) Vol. 188. | 1 | 9 | 0 |
| —— (B) Vol. 187. | 0 | 19 | 0 |
| 1897 (B) Vol. 188. | 0 | 19 | 6 |
| —— (A) Vol. 189. | 1 | 0 | 0 |
| —— (A) Vol. 190. | 1 | 8 | 6 |
| —— (B) Vol. 189. | 0 | 17 | 6 |
| 1898 (B) Vol. 190. | 2 | 5 | 0 |
| —— (A) Vol. 191. | 1 | 16 | 0 |
| 1899 (B) Vol. 191. | 1 | 12 | 0 |
| —— (A) Vol. 192. | 1 | 8 | 6 |
| 1900 (B) Vol. 192. | 1 | 7 | 6 |
| —— (A) Vol. 193. | 1 | 10 | 0 |
| —— (A) Vol. 194. | 1 | 1 | 6 |
| —— (B) Vol. 193. | 1 | 5 | 0 |
| 1901 (A) Vol. 195. | 1 | 7 | 0 |
| —— (A) Vol. 196. | 1 | 11 | 0 |
| —— (B) Vol. 194. | 3 | 0 | 0 |
| —— (A) Vol. 197. | 1 | 15 | 0 |
| 1902 (A) Vol. 198. | 1 | 8 | 6 |
| —— (A) Vol. 199. | 1 | 6 | 0 |
| 1903 (A) Vol. 200. | 2 | 3 | 0 |
| —— (A) Vol. 201. | 1 | 16 | 6 |
| —— (B) Vol. 195. | 1 | 7 | 0 |
| 1904 (A) Vol. 202 | 1 | 13 | 0 |
| —— (A) Vol. 203. | 1 | 1 | 6 |
| —— (B) Vol. 196. | 1 | 16 | 0 |
| 1905 (A) Vol. 204. | 1 | 9 | 0 |
| —— (B) Vol. 197. | 1 | 13 | 6 |
| 1906 (A) Vol. 205. | 1 | 8 | 0 |
| —— (A) Vol. 206. | 1 | 8 | 6 |
| —— (B) Vol. 198. | 2 | 1 | 0 |
| 1908 (A) Vol. 207. | 1 | 13 | 6 |
| —— (A) Vol. 208. | 1 | 18 | 6 |
| —— (B) Vol. 199. | 1 | 16 | 0 |
| 1909 (A) Vol. 209. | 1 | 13 | 0 |
| —— (B) Vol. 200. | 2 | 5 | 6 |
| 1910 (A) Vol. 210. | 1 | 3 | 6 |
| —— (B) Vol. 201. | 1 | 15 | 6 |
| 1912 (A) Vol. 211. | 1 | 6 | 6 |
| —— (B) Vol. 202. | 1 | 11 | 6 |
| 1913 (A) Vol. 212. | 1 | 3 | 6 |
| —— (B) Vol. 203. | 1 | 5 | 0 |
| 1914 (A) Vol. 213. | 1 | 6 | 0 |
| —— (A) Vol. 214. | 1 | 11 | 6 |
| —— (B) Vol. 204. | 1 | 12 | 6 |
| —— (B) Vol. 205. | 2 | 3 | 6 |
| 1915 (A) Vol. 215. | 1 | 11 | 0 |
| —— (B) Vol. 206. | 1 | 15 | 0 |
| 1916 (A) Vol. 216. | 1 | 10 | 0 |
| —— (B) Vol. 207. | 2 | 8 | 6 |
| 1918 (A) Vol. 217. | 1 | 11 | 0 |
| —— (B) Vol. 208. | 2 | 7 | 0 |
| 1919 (A) Vol. 218. | 1 | 15 | 0 |
| —— (A) Vol. 219. | 0 | 17 | 6 |
| 1920 (B) Vol. 209. | 3 | 13 | 0 |
| —— (A) Vol. 220. | 2 | 4 | 0 |
| 1921 (B) Vol. 210. | 3 | 5 | 6 |
| —— (A) Vol. 221. | 2 | 12 | 6 |

PUBLISHED BY HARRISON AND SONS, LTD., [illegible]ARTIN'S LANE, LONDON, W.C.